Seminar Paper

Complexity in Factor Pricing Models

Department of Statistics
Ludwig-Maximilians-Universität München

John-Pierre Weideman

Munich, August, 2024

Supervised by Prof. Daniel Wilhelm

Abstract

This paper investigates the impact of model complexity on factor pricing models, under the stochastic discount factor (SDF) framework. We build on the findings of Didisheim et al. (2024). We introduce and derive both the ordinary least squares and ridge estimators and clarifies the necessity of the ridge estimator in high-complexity scenarios. Through theoretical analysis, we demonstrate how complexity affects expected returns under different assumptions of the SDF model.

Contents

1 Introduction

This paper examines the the impact of complexity on factor pricing models, focusing particularly on the stochastic discount factor (SDF) framework. Our analysis is based on the work of Didisheim et al. (2024). The objective is to clarify some of their presented results.

Section 2 introduces the SDF model, which serves as the foundational model for this paper. Here, we discuss what an SDF model is, how it is used to price assets, and show when the resulting SDF prices assets correctly.

Section 3 presents a detailed derivation of both the OLS and ridge estimators for the SDF. We clarify why the need for regularized models arise in the case where model complexity is high. We demonstrate how the ridge estimator is used to overcome the shortcomings of the OLS estimator when the number of factors exceeds the number of observations.

In Section 4, we building on the theoretical framework of Didisheim et al. (2024), where we explain the differences in the behavior of expected returns between correctly specified and mis-specified SDF models. Didisheim et al. (2024) does not explicitly clarify why these differences arise. we replicate the theoretical graphs to provide a comparative analysis. Additionally, we critique a statement made by Didisheim et al. (2024) regarding the expected returns in the mis-specified case.

1.1 Literature Review

This paper discusses the ideas presented by Didisheim et al. (2024) in the paper "Complexity in Factor Pricing Models". Didisheim et al. (2024) looks at how model complexity affects the out-of-sample performance of SDF models. Their main contribution is to demonstrate, both theoretically and empirically, that higher complexity consistently enhances model performance. This challenges the econometric principle of parsimony (Tukey, 1961), where one prefer simple models that adequately captures the important features of the data, avoiding unnecessary complexity. The authors demonstrate that more complex models consistently outperform simpler ones. However, this only holds true for models that are properly regularised. Didisheim et al. (2024) adds to the literature on machine learning techniques to directly estimate the SDF from characteristic-based factors. Empirical studies on this topic include works by Brandt et al. (2009), Kozak et al. (2020) and Chen et al. (2023). Didisheim et al. (2024) adds to this literature by providing a theoretical framework for employing highly parameterized models for the estimation of the SDF.

"Complexity in Factor Pricing Models" also adds to the research on machine learning methods for analyzing factor pricing models, such as the works by Connor et al. (2012), Fan et al. (2016) and Kelly et al. (2020). Many researchers in this field argue that a small number of leading principal components can sufficiently explain the cross-section of returns, typically resulting in a low-complexity model environment where the true conditional SDF can be accurately estimated. In contrast, Didisheim et al. (2024) shows that, despite some drawbacks, highly complex models delivers better out-of-sample performance than low-complexity models.

The ideas in “Complexity in Factor Pricing” build on the earlier work, ”Virtue of Complexity in Return Prediction” by Kelly et al. (2021). This earlier paper looks at the benefits of using highly parameterised models for predicting returns. It shows that, in general, the performance of time series forecasting models gets better as they become more complex. However, the paper on complexity in factor models builds on the earlier paper in two ways. Firtly by switch the focus from time series forecasting to SDF estimation. And secondly, instead of looking at one asset in isolation, they now consider the panel setting with an arbitrary number of risky assets.

Didisheim et al. (2024), similar to Kelly et al. (2021) studies a class of high-dimensional ridge estimators. And in doing so, draws heavily on results of random matrix theory from Marenko and Pastur (1967).

2 Stochastic Discount Factor

The analysis of complexity in factor pricing models is built around the stochastic discount factor (SDF). A true SDF, if one exists, is representable as a tradable portfolio (Hansen and Richard, 1987):

$$M_{t+1} = 1 - w\left(X_t\right)' R_{t+1}, \tag{1}$$

$$s.t. \quad \mathbb{E}_t\left[R_{i,t+1} M_{t+1}\right] = 0, i = 1, ..., N \tag{2}$$

Where $R_{t+1} \in \mathbb{R}^{N \times 1}$ is the vector of excess returns on the $i = 1, ..., N$ risky assets. $w\left(X_t\right) \in \mathbb{R}^{N \times 1}$ is the vector containing the SDF's portfolio weights conditional on the information available until time t. X_t represents the conditioning variables, which can include macroeconomic indicators, market signals and other relevant financial data.

The SDF in (1) is a function that discounts future payoffs to their present value (see section 2.1). Condition (2) ensures that assets are priced correctly by ensuring that expected discounted excess returns are zero.

2.1 SDF properties

An SDF is used to price assets, by discounting future payoffs to present value P_{t+1} (Cochrane, 2001).

$$P_{i,t} = E_t[M_{t+1}(P_{i,t+1} + D_{i,t+1})]$$

Where:

$P_{i,t}$ is the price of the assets i in our portfolio, discounted to present value at time t.

$P_{i,t+1} + D_{i,t+1}$ is the future payoff (in the next period $t + 1$) of asset i.

An asset i is correctly priced if the expected discounted gross return of the asset equal 1.

$$\mathbb{E}_t\left[M_{t+1} R^g_{i,t+1}\right] = 1 \tag{3}$$

Where:

$R^g_{i,t+1} \equiv \frac{P_{i,t+1}+D_{i,t+1}}{P_{i,t}}$ is the future gross return.

This implies that the expected discounted excess return of our asset is:

$$\mathbb{E}_t\left[M_{t+1} R_{i,t+1}\right] = 0 \tag{4}$$

Where:

$R_{i,t+1} = R^g_{i,t+1} - R_f$ is the future excess return.

R_f is the risk-free rate.

The relationship between (3) and (4) holds under the assumption that there is no arbitrage opportunities in the market (Back, 2017). This means that the expected discounted payoff of a risk-free asset will equal its current price which is assumed to be 1, without loss of generality. Thus $\mathbb{E}_t\left[M_{t+1}\right]=\frac{1}{R_f}$, which leads to the result:

$$
\begin{aligned}
R_{i,t+1} &= R^g_{i,t+1} - R_{i,t} \\
\mathbb{E}_t\left[M_{t+1}R_{i,t+1}\right] &= \mathbb{E}_t\left[M_{t+1}R^g_{i,t+1}\right] - \mathbb{E}_t\left[M_{t+1}R_{i,t}\right] \\
\mathbb{E}_t\left[M_{t+1}R_{i,t+1}\right] &= 1 - \mathbb{E}_t\left[M_{t+1}\right]R_f \quad \text{(from (3))} \\
\mathbb{E}_t\left[M_{t+1}R_{i,t+1}\right] &= 1 - 1 \quad \text{(no-arbitrage condition)} \\
\mathbb{E}_t\left[M_{t+1}R_{i,t+1}\right] &= 0
\end{aligned}
$$

This result holds for $i=1,...,N$ for a market with N risky assets.

2.2 Complex SDF

The fundamental problem in cross-sectional asset pricing is the estimation of the weight vector $w\left(X_t\right)$. In the literature this estimation is often guided by the principle of parsimony, and thus focuses on tightly constrained functions of w. An example is the Fama and French (1993) three factor model. This is a model that restricts the SDF to a three parameter combination of factors consisting of the market risk, size factor and value factor [1].

However, in light of new machine learning models which often include a massive number of parameters, we can consider instead expanding the parameterization of the SDF to evaluate a complex SDF with an extremely large number of factors

Complexity in an SDF model is defined as $c=P/T$ Where P is the number of features (factors) and T is the number of training data points. A complex SDF is one where $P,T\to\infty$ and $c>1$. An SDF of this sort is a factor model with a very large amount of factors. This large SDF model can be approximated by

$$
w(X_t) \approx \sum_{p=1}^{P} \lambda_p S_p\left(X_t\right) \tag{5}
$$

where each $S_p\left(X_t\right)\in\mathbb{R}^{N\times 1}$ is a nonlinear basis function of X_t. Thus, the SDF is approximated by

$$
M_{t+1} \approx 1 - \sum_{p=1}^{P} \lambda_p S_p\left(X_t\right)' R_{t+1} = 1 - \lambda' F_{t+1} \tag{6}
$$

where $\lambda\in\mathbb{R}^{P\times 1}$, $F_{t+1}\in\mathbb{R}^{P\times 1}$, and each factor $F_{p,t+1}=S_p\left(X_t\right)'R_{t+1}$ is a characteristic-managed portfolio of base assets, which is constructed by using the non-linear "asset characteristics" $S_p\left(X_t\right)$ as portfolio weights.

The main result of Didisheim et al. (2024) shows that the out-of-sample pricing errors obtained from the SDF in (6) decreases as the number of factors increases and that the Sharpe ratio of the SDF portfolio increases with increased complexity.

[1] The three factors are represented by market excess return (MKT), out performance of small companies over big companies (SMB), and the out performance of high book-to-market compared to low book-to-market companies (HML). The coefficients assigned to each of the three factors are captured by the weight vector $w(X_t)=(w_{MKT},w_{SMB},w_{HML})'$. The returns on the MKT, SMB and HML portfolios are $R_{t+1}=(MKT,SMB,HML)'$. Which leads to the SDF $M_{t+1}=1-(w_{MKT}\cdot MKT_{t+1}+w_{SMB}\cdot SMB_{t+1}+w_{HML}\cdot HML_{t+1})$

2.2.1 Varying SDF Complexity

To vary the SDF's complexity, the machine learning method of random feature regression is used. This allows us to convert the fixed set of features X_t into any desired number P of random features, which subsequently allows for the evaluation of SDF complexity. Random feature regression transforms the input X_t into a higher-dimensional feature space S using random projections. This is achieved by generating random weights W and biases b and then applying a nonlinear activation function ϕ.

$$S(X_t) = \phi(WX_t + b)$$

Once the data is projected into the higher-dimensional space S, a linear regression model is fitted to the transformed data. The goal is to find the weight vector $\hat{\beta}$ that minimizes the error between the predicted values $\hat{y} = \hat{\beta}S$ and actual values y. In our case we have

$$\hat{w}(X_t) = \hat{\lambda}' S\left(X_t\right),$$

where $\hat{\lambda}$ is the vector we need to estimate in order to obtain the optimal weights $\hat{w}$ which is as close as possible to the w which provides the optimal mean-variance efficient portfolio. The next section clarifies how this estimator is obtained.

3 SDF Estimator

Portfolio selection is an important issue in finance which focuses on the effective allocation of investor resources to achieve optimal growth in investment. This issue is closely connected to many challenges in asset pricing. Under weak economic assumptions, such as the absence of arbitrage, the mean-variance efficient (MVE) portfolio within an economy represents a tradable form of the SDF. Hansen and Richard (1987) shows that the MVE portfolio is theoretically equivalent to the SDF. Thus, from this equivalence, the SDF can be seen as capturing the way investors balance risk and return to arrive at equilibrium prices.

The SDF is essential for both predicting asset prices and constructing optimal portfolios. It determines equilibrium prices through a mechanism that balances the risk and return of a portfolio of assets.

Markowitz (1952) introduced the concept of mean-variance optimization in portfolio theory, where investors construct portfolios that balance risk and return based on their individual preferences (represented by utility functions). The original Markowitz model assumes complete knowledge of the return distributions, which allows for the estimation of expected returns, variances, and covariances. However, in practice, investors must estimate these distributions with limited data, making the portfolio selection problem more complex.

Recent literature has proposed new solutions to this estimation challenge using machine learning (ML) techniques, such as shrinkage, model selection, and flexible parameterization. These ML methods for portfolio optimization focus on estimating portfolio rules while still adhering to the framework of utility optimization. A key contribution of modern ML approaches is their integration of utility maximization directly into the statistical problem of optimizing portfolio weights, without requiring any assumptions about the distribution of returns (Ait-Sahalia and Brandt., 2001, Brandt and Santa-Clara, 2006, Brandt et al., 2009). In these ML-based approaches, investor utility and the functional form of the portfolio weight function are defined based on observable covariates. The parameters of this portfolio weight function are determined by maximizing average utility within the sample, as described by Brandt et al. (1999).

In our estimation of the SDF parameters, we also focus on the mean-variance utility framework, consistent with Markowitz's approach. This framework is chosen because it allows the representation of the parametric portfolio weight using an OLS regression. Specifically, we refer to Theorem 1 from Britten-Jones (1999), which shows that the solution to the Markowitz optimization problem is proportional to the OLS coefficient obtained from regressing a constant (representing a risk-free return) on the excess returns R_{t+1}. This is equivalent to finding the combination of risky assets that yields the highest possible in-sample Sharpe ratio (Kelly and Xiu, 2023).

3.1 OLS SDF Estimator

From the motivation above, we use the OLS regression, given by,

$$1 = w'R_{t+1} + u_{t+1}.$$

Or equivalently by using the formulation for w defined in (5) and (6):

$$1 = \lambda' S' R_{t+1} + u_{t+1} = \lambda' F_{t+1} + u_{t+1} \tag{7}$$

Where u_{t+1} is the error term.
The resulting solution for the coefficients λ which minimizes the objective function is

$$\lambda = \mathbb{E}_t \left[F_{t+1} F_{t+1}'\right]^{-1} \mathbb{E}_t \left[F_{t+1}\right] \tag{8}$$

Where $\mathbb{E}_t \left[F_{t+1}\right]$ and $\mathbb{E}_t \left[F_{t+1} F_{t+1}'\right]$ are the (theoretical) expected value and covariance of the factors over the true data distribution.
The empirical counterpart is:

$$\hat{\lambda} = \hat{\mathbb{E}} \left[F_t F_t'\right]^{-1} \hat{\mathbb{E}} \left[F_t\right] \tag{9}$$

Where $\hat{\mathbb{E}}[F_t] = \frac{1}{T} \sum_t F_t$ and $\hat{\mathbb{E}}[F_t F_t'] = \frac{1}{T} \sum_t F_t F_t'$ are the sample mean and covariance of the factors.

3.1.1 Derivation of OLS SDF estimator

To derive (8) and (9), we again consider the OLS regression for the SDF given in (7) by

$$1 = \lambda' F_{t+1} + u_{t+1}.$$

The theoretical MSE is defined as the expected value (over the true population data distribution) of the squared difference between the actual value (1 in this case) and the predicted values $\lambda' F_{t+1}$:

$$\text{MSE} = \mathbb{E}_t[(1 - \lambda' F_{t+1})^2]$$

Expanding the square:

$$\begin{aligned} \text{MSE} &= \mathbb{E}_t \left[(1 - 2\lambda' F_{t+1} + \lambda' F_{t+1} F_{t+1}' \lambda)\right] \\ &= 1 - 2\lambda' \mathbb{E}_t[F_{t+1}] + \lambda' \mathbb{E}_t[F_{t+1} F_{t+1}'] \lambda \end{aligned}$$

To minimize the theoretical MSE, we differentiate the MSE with respect to λ and set the derivative equal to zero:

$$\frac{\partial}{\partial \lambda} \text{MSE} = -2\mathbb{E}_t[F_{t+1}] + 2\mathbb{E}_t[F_{t+1} F_{t+1}'] \lambda = 0$$

Which gives us the solution for the optimal λ:

$$\lambda = \left(\mathbb{E}_t[F_{t+1} F_{t+1}']\right)^{-1} \mathbb{E}_t[F_{t+1}] \quad \text{(assuming } \mathbb{E}_t \left[F_{t+1} F_{t+1}'\right] \text{ is non-singular)}$$

Similarly, from the empirical MSE

$$\hat{\text{MSE}} = \frac{1}{T} \sum_t [(1 - \lambda' F_t)^2] = \hat{\mathbb{E}}[(1 - \lambda' F_t)^2],$$

we find the OLS estimator for λ, based on a finite sample of data:

$$\hat{\lambda} = \hat{\mathbb{E}} \left[F_t F_t'\right]^{-1} \hat{\mathbb{E}} \left[F_t\right] \quad \text{(assuming } \hat{\mathbb{E}} \left[F_t F_t'\right] \text{ is non-singular)}$$

The empirical estimator is simply the sample version of the theoretical estimator, replacing population expectations with sample averages.

3.2 Ridge SDF Estimator

Consider now the SDF estimator (9) in the high complexity case where $P \to \infty$ and $P > T$. In the high complexity case $\hat{\mathbb{E}}[F_t F_t']$ is singular and therefore not invertible. Thus, there is no unique solution since we have a system of equations which is underdetermined, meaning there are more unknowns than equations. In this case, the OLS estimator $\hat{\lambda}$ has an infinite number of solutions, all of which exactly fit the training data. This deficiency of $\hat{\mathbb{E}}[F_t F_t']$ is overcome by introducing a ridge penalty into the regression problem. Which leads to the ridge solution:

$$\begin{aligned}\hat{\lambda}(z) &= \arg\min_{\lambda} \left\{ \sum_{t=1}^{T} (1 - \lambda' F_t)^2 + z||\lambda||^2 \right\} \\ &= \left(zI + \hat{\mathbb{E}}[F_t F_t'] \right)^{-1} \hat{\mathbb{E}}[F_t] \end{aligned} \tag{10}$$

The SDF solution $\hat{\lambda}(z)$ is unique, well defined and has a finite in-sample Sharpe ratio, even in the high complexity case. The corresponding ridge portfolio return is

$$\hat{R}^M_{T+1}(z; P; T) = \hat{\lambda}(z)' F_{T+1}, \tag{11}$$

and SDF is

$$\hat{M}_{T+1}(z; P; T) = 1 - \hat{R}^M_{T+1}(z; P; T). \tag{12}$$

3.2.1 Derivation of the Ridge SDF estimator

To derive (10), we need the empirical Mean Squared Error (MSE) with ridge regularization, which is given by:

$$\begin{aligned}\hat{\text{MSE}}_r &= \frac{1}{T} \sum_t (1 - \lambda' F_t)^2 + z \sum_p \lambda_p^2 \\ &= \hat{\mathbb{E}}[(1 - \lambda' F)^2] + z\lambda'\lambda \end{aligned}$$

Expanding the MSE term:

$$\begin{aligned}\hat{\text{MSE}}_r &= \hat{\mathbb{E}}[1 - 2\lambda' F_t + \lambda' F_t F_t' \lambda] + z\lambda'\lambda \\ &= 1 - 2\lambda'\hat{\mathbb{E}}[F_t] + \lambda'\hat{\mathbb{E}}[F_t F_t']\lambda + z\lambda'\lambda \end{aligned}$$

Next, we take the derivative of the Ridge MSE with respect to λ and set it to zero in order to find the λ that minimizes the ridge MSE:

$$\frac{\partial}{\partial \lambda} \hat{\text{MSE}}_r = -2\hat{\mathbb{E}}[F_t] + 2\hat{\mathbb{E}}[F_t F_t']\hat{\lambda}(z) + 2z\hat{\lambda}(z) = 0$$

And solving in terms of $\hat{\lambda}(z)$ to find the ridge estimator:

$$\begin{aligned}(\hat{\mathbb{E}}[F_t F_t'] + zI)\hat{\lambda}(z) &= \hat{\mathbb{E}}[F_t] \\ \hat{\lambda}(z) &= (\hat{\mathbb{E}}[F_t F_t'] + zI)^{-1}\hat{\mathbb{E}}[F_t] \end{aligned}$$

3.2.2 Motivation for using Ridge in the High Complexity regime

Consider the system of equations arising when determining the OLS estimator:

$$\hat{\mathbb{E}}[F_t F_t']\hat{\lambda} = \hat{\mathbb{E}}[F_t]$$

In high-complexity scenarios where $(P > T)$, the matrix $\hat{\mathbb{E}}[F_t F_t']$ is not of full rank and is singular. This means that $\hat{\mathbb{E}}[F_t F_t']$ is not invertible, and the system of equations do not have a unique solution. Ridge regression introduces a penalty term $z > 0$ to address the problem of singularity and stabilize the solution, resulting in the modified system:

$$\left(\hat{\mathbb{E}}[F_t F_t'] + zI\right)\hat{\lambda}(z) = \hat{\mathbb{E}}[F_t]$$

where $I \in \mathbb{R}^{P \times P}$ the identity matrix. The addition of zI ensures that the matrix $\hat{E}_t[F_t F_t'] + zI$ is non-singular and thus invertible. The matrix $\hat{E}_t[F_t F_t']$ is symmetric and positive semi-definite, meaning all its eigenvalues are non-negative. However, in the case where $P > T$, some of these eigenvalues could be zero, which causes $\hat{E}_t[F_t F_t']$ to be singular. By adding zI, we effectively shift all eigenvalues of $\hat{E}_t[F_t F_t']$ by z. Since $z > 0$, even if some eigenvalues of $\hat{E}_t[F_t F_t']$ were zero, the eigenvalues of $\hat{E}_t[F_t F_t'] + zI$ become strictly positive, ensuring that the matrix $\hat{E}_t[F_t F_t'] + zI$ becomes positive definite and thus invertible.

4 Expected Returns of the Complex SDF

In this section we compare the properties of the complex SDF in two distinct cases. The one being the correctly specified SDF where we assume that the factors F_t represents the true and complete set of factors. The other case will consider the mis-specified SDF model. The mis-specified case is more realistic for practical purposes, because in reality one does not know the true and complete set of factors. Instead, it is assumed that only a fraction $q = \frac{P_1}{P} < 1$ of the factors are observable. The observable factors are represented by $F_t(q) = (F_{i,t})_{i=1}^{P_1}$. And we are interested in the properties of the SDF as P_1 increases and approaches the true number of factors P.
Firstly, for the setting of the **correctly specified model**, we will need the ridge SDF estimator defined in (10) as,

$$\hat{\lambda}(z) = \hat{\lambda}(z; P; T) = \left(zI + \hat{E}[F_t F_t']\right)^{-1} \hat{E}[F_t],$$

where $\hat{E}[F_t] = \frac{1}{T}\sum_t F_t$ and $\hat{E}[F_t F_t'] = \frac{1}{T}\sum_t F_t F_t'$ are again the sample mean and covariance of factors. The corresponding ridge portfolio return and SDF are

$$\hat{R}^M_{T+1}(z; P; T) = \hat{\lambda}(z)' F_{T+1}, \quad \hat{M}_{T+1}(z; P; T) = 1 - \hat{R}^M_{T+1}(z; P; T).$$

The corresponding infeasible SDF estimator (which is the theoretical true model) is

$$\lambda(z) = (zI + \mathbb{E}[FF'])^{-1}\,\mathbb{E}[F] \tag{13}$$

and its return and SDF,

$$R^M_T(z) = \lambda(z)' F_{T+1}, \quad M_{T+1}(z) = 1 - R^M_T(z).$$

We define the following formulas for our analysis:

$$\mathcal{E}(z) \equiv \mathbb{E}\left[R^M_{T+1}(z)\right] = \mathbb{E}[F]'(zI + \mathbb{E}[FF'])^{-1}\mathbb{E}[F] \in (0, 1), \tag{14}$$

Similarly, for the **mis-specified model** we define the SDF estimates as

$$\hat{\lambda}(z; P_1; T) = \left(zI + \hat{E}[F_t(q)F_t(q)']\right)^{-1} \hat{E}[F_t(q)]$$

with corresponding portfolio return and SDF

$$\hat{R}^M_T(z; P_1; T) = \hat{\lambda}(z; P_1; T)' F_{T+1}(q), \quad \hat{M}_{T+1}(z; P_1; T) = 1 - \hat{R}^M_T(z; P_1; T).$$

We also define the analog of the infeasible ridge SDF estimator (13) for the mis-specified case:

$$\lambda(z; q) = (zI + \mathbb{E}[F(q)F(q)'])^{-1}\,\mathbb{E}[F(q)] \quad (59)$$

and its return and SDF,

$$R^M_T(z; q) = \lambda(z; q)' F_{T+1}, \quad M_{T+1}(z; q) = 1 - R^M_T(z; q).$$

The analog of functions (14) for the mis-specified model which depends on q is

$$\mathcal{E}(z;q) = \mathbb{E}[R_T^M(z;q)] = \mathbb{E}[F(q)]'(zI + \mathbb{E}[F(q)F(q)'])^{-1}\mathbb{E}[F(q)]. \tag{15}$$

4.1 Expected Returns in the Correctly Specified Model

The expected value of the out-of-sample ridge SDF return is derived as:

$$\begin{aligned}
\mathbb{E}[\hat{R}^M_{T+1}(z;P;T)] &= \mathbb{E}\left[\hat{\lambda}(z)'F_{T+1}\right] \quad \text{(from (11))} \\
&= \mathbb{E}\left[\hat{\mathbb{E}}[F_t]'\left(zI + \hat{\mathbb{E}}[F_tF_t']\right)^{-1}F_{T+1}\right] \\
&= \left(\frac{1}{T}\sum_t \mathbb{E}[F_t]\right)'\left(zI + \frac{1}{T}\sum_t \mathbb{E}[F_tF_t']\right)^{-1}\mathbb{E}[F_{T+1}] \quad \text{(sample mean and covariance)} \\
&= \left(\frac{1}{T}\sum_t \mathbb{E}[F]\right)'\left(zI + \frac{1}{T}\sum_t \mathbb{E}[FF']\right)^{-1}\mathbb{E}[F] \quad (\mathbb{E}[F_t] = \mathbb{E}[F]\ \forall t) \\
&= \frac{1}{T}\cdot T(\mathbb{E}[F])'\left(zI + \frac{1}{T}\cdot T(\mathbb{E}[FF'])\right)^{-1}\mathbb{E}[F] \quad (\sum_t c = T\cdot c) \\
&= \mathbb{E}[F]'\left(zI + \mathbb{E}[FF']\right)^{-1}\mathbb{E}[F]
\end{aligned} \tag{16}$$

Didisheim et al. (2024) proves, that *in the limit*, as $P, T \to \infty$, $P/T \to c$, the expected out-of-sample return of the ridge SDF satisfies

$$\lim \mathbb{E}[\hat{R}^M_{T+1}(z;P;T)] = \mathcal{E}(Z^*(z;c)),$$

where $Z^*(z;c)$ is given by

$$Z^*(z;c) = z(1+\xi(z;c)) \in (z, z+c), \tag{17}$$

and $\xi(z;c)$ is defined by

$$\xi(z;c) = \frac{c(1-m(-z;c)z)}{1-c(1-m(-z;c)z)}{}^{2} \in (0, \frac{c}{z})^{3}$$

Thus, the the expected value in the limit is

$$\lim \mathbb{E}[\hat{R}^M_{T+1}(z;P;T)] = \mathbb{E}[F]'(Z^*(z;c)I + \mathbb{E}[FF'])^{-1}\mathbb{E}[F], \tag{18}$$

[2]
$$m(-z;c) = \lim_{P\to\infty, P/T\to c} \frac{1}{P}\mathrm{tr}\left(\left(zI + \hat{E}[F_tF_t']\right)^{-1}\right),$$

$m(-z,c)$ describes the Stieltjes transform of the empirical eigenvalue distribution of the factor covariance matrix in the high-complexity limit. The Stieltjes transform is used in random matrix theory to linking the distribution of eigenvalues to the behavior of large-dimensional matrices

[3] To find the range of $\xi(z;c)$, we start with the inequality

$$z(1+\xi(z;c)) \in (z, z+c).$$

which differs from the non-limiting case in (18), where the standard shrinkage term z is replaced by Z^*. The result of expected returns in the limit is given as a theorem by Didisheim et al. (2024) where they show how Z^* enters the equation. Furthermore, they state (without proof) that Z^* is monotonically increasing in z and in c. For the discussion we assume the statement of monotonicity of Z^* to be correct.

From the definition of (17) we can see the additional *implicit shrinkage* that complexity introduces. When complexity is 0, $z = Z^*(z, 0)$, and the explicit shrinkage is the only shrinkage present. But when $c > 0$, we see that $Z^* > z$, because of implicit shrinkage generated by complexity. When holding the ridge parameter z constant, and increasing the number of parameters P, the norm $\|\lambda\|^2$ of the coefficient vector cannot increase due to the ridge penalty, which constrains the overall size of the coefficients. As more parameters are added to the model, to satisfy the ridge constraint, the coefficients in the vector λ must shrink further. This shrinkage is considered "implicit" because it results naturally from the ridge penalty's properties rather than being explicitly imposed by increasing the explicit ridge parameter z. The increase in the number of parameters forces a greater shrinkage on the coefficients to maintain the penalty constraint.

From (18) we can determine the effect of increasing complexity on expected returns. Under the correctly specified model, $\mathbb{E}[F]$ and $\mathbb{E}[FF']$ remains constant because F is the true and complete set of factors. However, as complexity c increases, the term $Z^*(z, c)$ increases, which means that the inverse matrix $(Z^*(z; c)I + \mathbb{E}[FF'])^{-1}$ causes the expected returns to be decreasing in complexity. In figure 1, we demonstrate the theoretical behaviour of expected returns as complexity increases. The expected return consistently decreases in complexity for all values of the ridge penalty parameter. It is also clear that for lower levels of explicit regularization z, the expected returns are highest. We also observe stronger implicit regularization effects for less regularized models as complexity increases (depicted by more dramatic decreases in expected returns).

Which implies

$$
\begin{aligned}
&z < z(1 + \xi(z; c)) < z + c. \\
&\frac{z}{z} < \frac{z(1 + \xi(z; c))}{z} < \frac{z + c}{z}, \\
&1 < 1 + \xi(z; c) < 1 + \frac{c}{z}. \\
&0 < \xi(z; c) < \frac{c}{z}.
\end{aligned}
$$

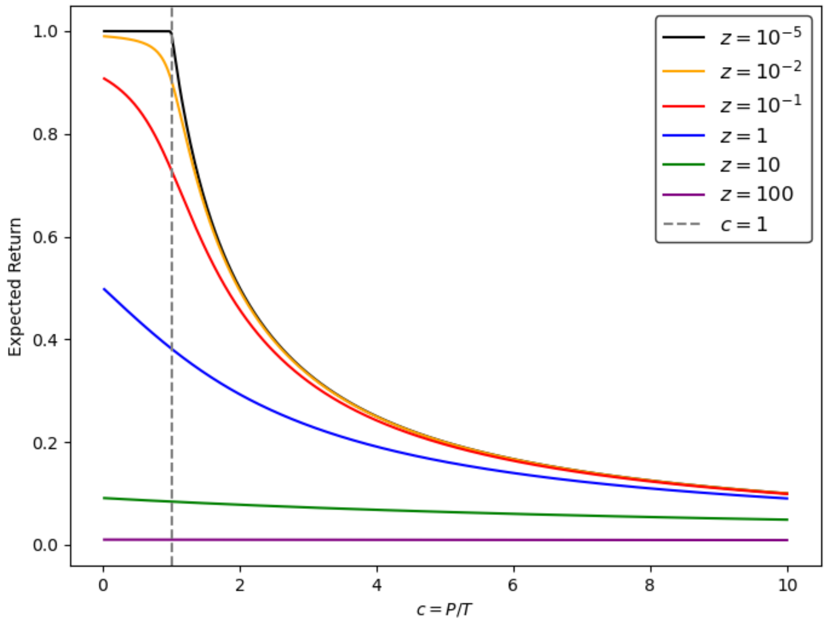

Figure 1: Expected return for Correctly Specified Model

4.2 Expected Returns in the Mis-Specified Model

With similar reasoning as in the correctly specified model, and from a theorem by Didisheim et al. (2024), we can evaluate the expected out-of-sample ridge SDF return for the mis-specified model. In the limit as $P_1, T \to \infty$, $P_1/T \to cq$,

$$\begin{aligned}\lim \mathbb{E}[\hat{R}^M_{T+1}(z; P_1; T)] &= \mathcal{E}(Z^*(z; cq; q); q) \\ &= \mathbb{E}[F(q)]'(Z^*(z; cq; q)I + \mathbb{E}[F(q)F(q)'])^{-1}\mathbb{E}[F(q)] \end{aligned} \tag{19}$$

where $Z^*(z; cq; q)$ is given by

$$Z^*(z; cq; q) = z\,(1 + \xi(z; cq; q)) \in (z, z + cq).$$ [4]

To understand the behaviour of expected returns as complexity increases, we evaluate the effect of changes in our complexity parameter $q = \frac{P_1}{P}$ where $cq = \frac{P_1}{T} \in (0, c)$ is the complexity of the mis-specified model.

$Z^*(z; qc; q)$ represents the effective shrinkage that is applied to the model. Similar to the correctly specified model, this term is influenced by both the explicit regularization parameter z and the complexity ratio cq. The effect of Z^* in the mis-specified model

4

$$\xi(z; cq) = \frac{cq\,(1 - m(-z; cq; q)z)}{1 - cq\,(1 - m(-z; cq; q)z)}$$

$$m(-z; cq; q) = \lim_{P_1\to\infty,\; P_1/T\to cq} \frac{1}{P_1}\,\mathrm{tr}\left(\left(zI + \hat{E}\left[F_t(q)F_t(q)'\right]\right)^{-1}\right)$$

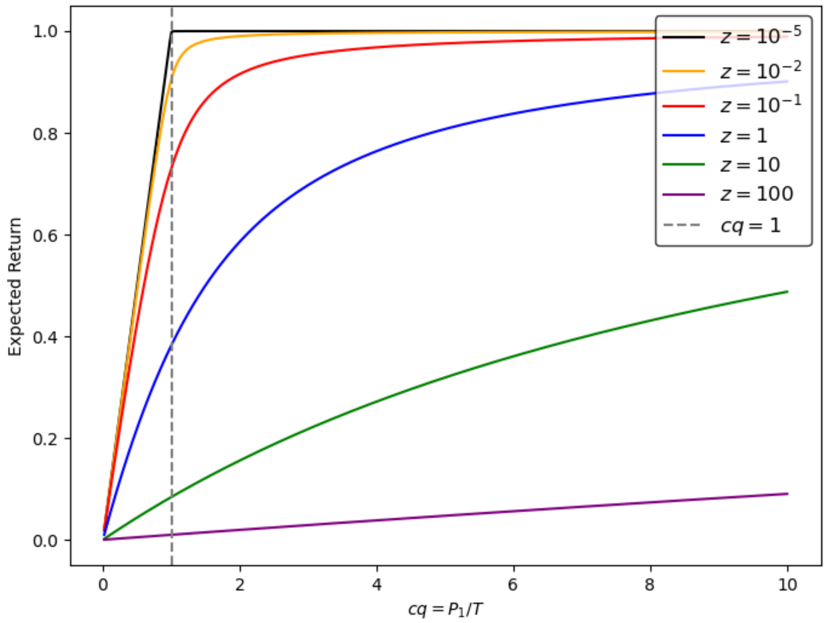

Figure 2: Expected Out-of-sample Return for Mis-specified Model

on expected returns is the same as in the correctly specified regime. As the model's complexity increases (with larger P_1), the value of Z^* increases, leading to stronger implicit shrinkage. This decreases the expected return.

However, there arises a distinction between the correctly specified and mis-specified model from the behaviour of our factor terms. In the mis-specified model, $F(q)$ represents the subset of the true (unknown) factors F, used in the model. As q increases, more factors are included in $F(q)$, and this potentially improves the model's approximation to the true SDF. Intuitively one would thus expect expected returns to be increasing in q, since higher complexity allows the model to approximate the true SDF more accurately. For this to be the case, the effect of increasing complexity should become clear by evaluating (19). As q increases, more factors are included in $F(q)$, potentially increasing $\mathbb{E}[F(q)]$ as more relevant information is included. However, increasing q will also expand the covariance matrix $\mathbb{E}[F(q)F(q)']$. And thus it is not straightforward to determine the overall impact of an increase in q on expected returns. Expected returns will be increasing in q if $\mathbb{E}[F(q)]'\mathbb{E}[F(q)]$ increases faster than the inverse matrix. However, Didisheim et al. (2024) asserts that out-of-sample expected return of the SDF portfolio is *always* increasing in model complexity. They derive this result from a theoretical simulation of expected returns, where they assume that $\mathbb{E}[F(q)]'\mathbb{E}[F(q)]$ increases proportionally with model complexity, while the covariance matrix remains an identiy matrix as q increases. This assumption leads to the behaviour of expected returns displayed in figure 2. Figure 2 is a replication of the "Virtue of Complexity" graphs by Didisheim et al. (2024). Under a specific set of assumptions it is true that expected returns are always increasing in complexity. However, we argue that contrary to the assertion of Didisheim et al. (2024) that

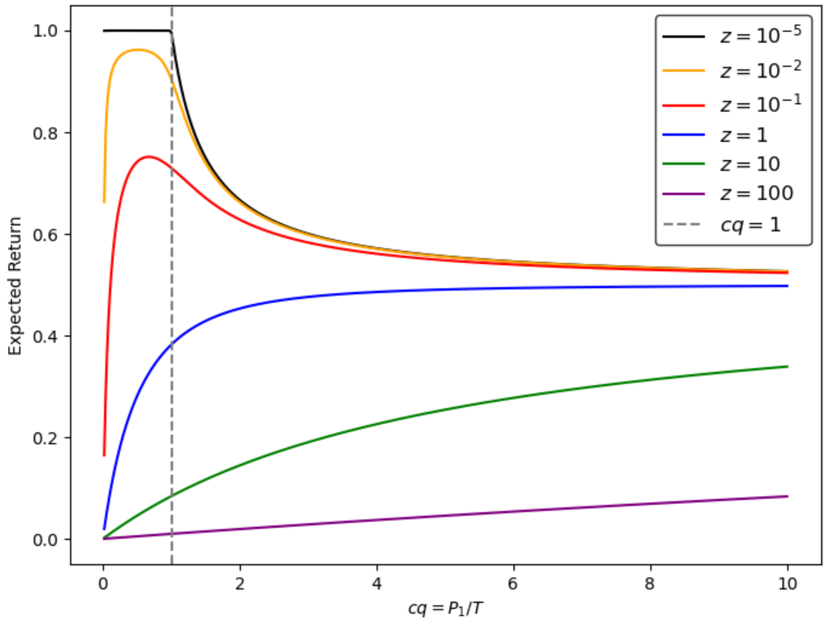

Figure 3: Expected Out-of-sample Return for Mis-specified Model under Varied Assumptions

expected returns are always increasing with added complexity, the behaviour of expected returns instead depends on the theoretical assumptions made on the terms in (19). If we, for example, drop the assumption of an identity matrix for the covariance structure, we see in figure 3 that the expected returns behave differently. If the covariance matrix is assumed to also vary with complexity, then for some levels of explicit regularization, the expected returns are increasing when complexity is low, but becomes decreasing with higher complexity. It is thus clear that the statement by Didisheim et al. (2024) that expected returns are always increasing in complexity does not hold. Instead, the behaviour is dependent on the theoretical assumptions on $F(q)$.

5 Conclusion

We explore the impact of model complexity on factor pricing models within the stochastic discount factor framework. We began by introducing the SDF model and discussed its role in asset pricing. Furthermore, we define the SDF framework under the high complexity regime. We derived both the ordinary least squares and ridge estimators and demonstrated the necessity of using regularized models like the ridge in high-complexity settings to overcome the limitations of the OLS estimator.

Our analysis extended the findings of Didisheim et al. (2024) by critically examining the differences in expected returns between correctly specified and mis-specified SDF models. We argued that contrary to the claims made by Didisheim et al. (2024), the behavior of expected returns is not universally increasing with added complexity, but is instead dependent on the specific theoretical assumptions on the model's factors. Our replication of graphs from Didisheim et al. (2024) and critique, supported by simulated results, showed that the assumptions about the covariance structure can lead to varying behaviour of expected returns in complex models.

A Appendix

The code used to generate the graphs presented in this paper can be found here: https://github.com/jpweideman/Complexity-in-Factor-Pricing. This repository includes the necessary scripts to replicate the results discussed in the paper.

References

Ait-Sahalia, Y. and Brandt., M. W. (2001). Variable selection for portfolio choice, *The Journal of Finance, Volume 56, pp. 1297–1351* .

Back, K. E. (2017). *Asset Pricing and Portfolio Choice Theory*, Oxford University Press.

Brandt, M. W. and Santa-Clara, P. (2006). Dynamic portfolio selection by augmenting the asset space, *The Journal of Finance, Vol 61, Issue 5, pp. 2187–2217* .

Brandt, M. W., Santa-Clara, P. and Valkanov, R. (1999). Estimating portfolio and consumption choice: A conditional euler equations approach, *The Journal of Finance, Volume 54, Issue 5, pp. 1609–1645* .

Brandt, M. W., Santa-Clara, P. and Valkanov, R. (2009). Covariance regularization by parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns, *Review of Financial Studies, Volume 22, pp. 3411–3447* .

Britten-Jones, M. (1999). The sampling error in estimates of mean- variance efficient portfolio weights, *The Journal of Finance, Volume 54, Issue 2, pp. 655–671* .

Chen, L., Pelger, M. and Zhu, J. (2023). Deep learning in asset pricing, *Management Science* .

Cochrane, J. H. (2001). *Asset Pricing*, Princeton University Press.

Connor, G., Hagmann, M. and Linton, O. (2012). Efficient semiparametric estimation of the fama–french model and extensions, *Econometrica, Volume 80, Issue 2, pp. 713-754* .

Didisheim, A., Ke, S., Kelly, B. and Malamud, S. (2024). Complexity in factor pricing models, *Swiss Finance Institute Research Paper No. 23-19* .

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics, Volume 33, Issue 1, Pages 3-56* .

Fan, J., Liao, Y. and Wang, W. (2016). Projected principal component analysis in factor models, *Annals of statistics, Volume 44, Issue 1, pp. 219* .

Hansen, L. P. and Richard, S. F. (1987). The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica, Vol. 55, No. 3, pp. 587-613.* .

Kelly, B. T., Malamud, S. and Zhou, K. (2021). The virtue of complexity in return prediction, *Swiss Finance Institute Research Paper* .

Kelly, B. T., Pruitt, S. and Su, Y. (2020). Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* .

Kelly, B. and Xiu, D. (2023). Financial machine learning, *Working Paper* .

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

Marenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices, *Mathematics of the USSR-Sbornik, Volume 1, Issue 457* .

Markowitz, H. (1952). Portfolio selection, *The Journal of Finance, Volume 7, No. 1, pp. 77-91* .

Tukey, J. W. (1961). Discussion, emphasizing the connection between analysis of variance and spectrum analysis, *Technometrics, Volume 3, Issue 2, pp. 191–219* .