Nonlinear Asset Covariance Prediction Conditioned on the Financial Environment

Alexander Hansson

Master's thesis
2026:E101

Lund Institute of Technology
Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

Master's Theses in Mathematical Sciences 2026:E101
ISSN 1404-6342

LUTFMS-3574-2026

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lu.se/

Abstract

The thesis studies the problem of predicting large covariance matrices, a central challenge in modern portfolio theory. In portfolios with many assets, the large number of covariance terms makes prediction difficult, often causing traditional methods to suffer from either substantial estimation bias or variance. To address this, the thesis proposes an approach that adapts conditional autoencoders, originally developed in asset pricing theory. The framework decomposes asset movements nonlinearly into a diagonal idiosyncratic component and a low-dimensional systematic component, conditionally on the financial environment. These are then forecast separately using dynamic covariance models. This setup enables nonlinear latent factor construction and, unlike traditional autoencoder approaches, allows the covariance structure to adapt to the changing financial environment through time-varying factor loadings.

Out-of-sample performance was evaluated on large-capitalization U.S. equities, using a value-weighted portfolio and minimum-variance portfolios constructed using covariance estimates from the proposed framework, a Ledoit-Wolf shrinkage estimator, and a traditional autoencoder. The models were compared using the Sharpe ratio and resulting transaction costs. Furthermore, each model's forecasting accuracy was evaluated against a target covariance matrix constructed from out-of-sample high-frequency return data. Results show that the proposed framework consistently outperforms the benchmarks across both statistical and economic metrics, most notably a higher Sharpe ratio for the minimum-variance portfolio. While the proposed framework leads to higher portfolio turnover, the net Sharpe ratio remains higher than the benchmarks under realistic transaction costs and weight restrictions. Overall, the results suggest that conditional autoencoders are a promising framework for high-dimensional covariance forecasting.

Acknowledgments

I would like to express my sincere gratitude to the Third Swedish National Pension Fund for its support and guidance throughout the course of this thesis. In particular, I would like to thank my supervisor at the pension fund, Gustav Fyring, for his valuable advice and insights. I am also grateful to Mervan Kaya for many helpful discussions on applying machine learning models in finance. Finally, I would like to thank Magnus Wiktorsson, my supervisor at the Centre for Mathematical Sciences, for his continuous support, insightful feedback, and contagious enthusiasm for the subject.

Nomenclature

Factor modeling

Σ_X	Covariance matrix of variable X
F_t	Latent factor vector
Z_t	Firm characteristic matrix
B_t	Factor loading matrix
X_t	Characteristic-managed portfolio returns
N	Number of assets
K	Number of latent factors
P	Number of firm characteristics

Finance

R_t	Asset return vector
u_t	Idiosyncratic return vector
μ	Mean return vector
σ	Standard deviation or volatility
w_t	Portfolio weight vector at time t

Mathematics

$\mathbb{R}^{X \times Y}$	Space of $X \times Y$ real-valued matrices
$\|\cdot\|_2$	Euclidean norm
$\|\cdot\|_F$	Frobenius norm
tr	Trace operator
diag	Diagonal matrix operator
∇	Gradient operator

Abbreviations

AE	Autoencoder
CAE	Conditional Autoencoder
VW	Value-weighted / market-capitalization-weighted
LW	Ledoit-Wolf shrinkage estimator
DCC	Dynamic Conditional Correlation
GARCH	Generalized Autoregressive Conditional Heteroskedasticity

Contents

Chapter 1

Background

1.1 Introduction

Every day, asset managers, pension funds, and banks allocate vast amounts of capital across financial assets, aiming to construct a portfolio that delivers high returns while maintaining minimal risk. Central to this process is diversification. By combining investments whose returns do not move together, the risk of significant losses can be reduced without sacrificing expected return. Achieving this requires an understanding of how assets move relative to one another, commonly represented by a covariance matrix.

While accurate covariance matrices are crucial for financial institutions, producing accurate estimates remains challenging. Firstly, financial portfolios are often high-dimensional, comprising hundreds or even thousands of individual assets whose pairwise relationships must be estimated simultaneously. Since the number of terms in the covariance matrix increases quadratically, reliable estimation quickly becomes challenging. For instance, a portfolio of 500 assets requires estimating 125,250 unique terms. Furthermore, the covariance structure constantly fluctuates. Periods of financial stress, policy changes, and evolving macroeconomic conditions may quickly alter the relationships among assets' relative movements (Sandoval Junior & Franca, 2012). Consequently, old observations are less representative of the current market environment. In addition, numerous terms must be estimated from a limited number of observations. The combination of high dimensionality and limited samples lays the foundation for a crucial yet challenging estimation task that can lead to substantial error if not handled well.

To alleviate the challenges of high-dimensional estimation, dimensionality reduction techniques can be used to compress the covariance structure into a low-dimensional space of so-called factors that drive covariance. The covariance is then estimated from the

factors and subsequently reconstructed to the original asset dimension. Recent literature has shown that nonlinear dimensionality reduction techniques provide significant improvements to covariance estimation compared to linear techniques (Conlon et al., 2025). Traditional autoencoder approaches, while specifically designed to perform nonlinear dimensionality reduction, sacrifice capacity to adapt to the changing financial environment. This is problematic, as covariance matrices are time-varying, especially across different economic regimes (Sandoval Junior & Franca, 2012). More specifically, autoencoders encode asset returns without conditioning on economic information and assume a static relationship between the covariance matrix and the underlying factors. These have proven to be fundamental limitations, as conditioning on the financial environment by allowing the relationship to be time-varying has been shown to yield superior covariance estimates for portfolio optimization relative to sophisticated benchmarks that do not. See, for instance, Fan et al. (2024) and Jung (2019).

To address the issues of autoencoders, the thesis proposes a framework for covariance estimation using the conditional autoencoder architecture of Gu et al. (2021). The architecture was originally proposed for asset pricing theory, but the framework adapts it to estimate covariance matrices. The framework preserves the nonlinear dimensionality reduction of traditional autoencoders while adapting the encoding to the financial environment. This is achieved by allowing the relationship between the covariance and factors to vary with firm characteristics.

1.1.1 Objective

The thesis aims to contribute to the literature on dimensionality-reduction techniques for estimating covariance matrices. It investigates whether nonlinear dimensionality reduction conditioned on the firm's characteristics can accurately produce time-varying covariance estimates, which serve as a basis for forecasting when combined with dynamic covariance models. To achieve this, it develops a new framework using conditional autoencoders and evaluates it based on statistical predictive accuracy and out-of-sample portfolio performance. This objective is investigated through the following research questions:

1. Can a conditional autoencoder framework improve covariance forecasting from both a statistical and portfolio performance perspective, relative to classical and existing autoencoder-based approaches?

2. Do the possible improvements achieved by the framework remain under realistic portfolio construction settings, including transaction costs and portfolio constraints?

1.1.2 Delimitations

Several delimitations apply to the thesis's scope. As the main focus is on constructing an architecture for covariance estimation using the conditional autoencoder, relatively unsophisticated forecasting methods are utilized, namely GARCH and DCC. These serve to demonstrate their integration into the architecture and their contribution to performance. Regarding the data, the analysis is restricted to U.S. large-cap equities that are current or former constituents of the S&P 500. Furthermore, all returns and economic data utilized by the models are monthly. From a methodology perspective, no expected return forecasting is performed. Rather, in a portfolio-construction setting, the covariance matrices are evaluated only by their resulting minimum-variance portfolio.

1.2 Thesis Structure

Chapters 2 through 4 constitute the theoretical foundation of the thesis. Chapter 2 introduces portfolio optimization, with particular emphasis on minimum-variance portfolio optimization. Furthermore, it discusses the challenges and various methods for covariance estimation, leading up to the core idea of factor models. Neural networks are introduced in Chapter 3, with a focus on autoencoders. This basis is subsequently used to review the autoencoder's role in covariance estimation. Chapter 4 presents the proposed conditional autoencoder framework, where dimensionality reduction of asset returns is performed conditioned on firm characteristics. It describes how the conditional autoencoder can be used within a framework to forecast covariance matrices and outlines metrics for evaluating the proposed framework.

Chapter 5 describes the methodology and data, including the datasets, preprocessing procedures, model training, portfolio construction, and evaluation approach. The results are then presented in Chapter 6, including comparisons against several strong benchmarks. It evaluates both the economic and statistical performance of the models. Lastly, Chapter 7 discusses the empirical findings, including the strengths and limitations of the proposed framework.

Chapter 2

Covariance Estimation for Portfolio Optimization

Markowitz (1952) introduced the mean-variance problem, which established the foundation for modern portfolio theory. The idea is that the expected returns and covariance matrix of returns can be used to compute a portfolio allocation that optimally balances return and risk. In the simplest case, these are estimated using sample moments, which often vary substantially over time and perform poorly out-of-sample (Broadie, 1993; Kan & Zhou, 2007) . This thesis focuses solely on the covariance portion of this problem. Under the assumption of identical and constant expected returns, the problem is reduced to constructing an optimal minimum-variance portfolio.

Let N denote the number of assets. The minimum-variance portfolio is obtained by finding the portfolio weights, $w \in \mathbb{R}^N$, that minimize the portfolio variance, given an estimated covariance matrix, $\Sigma_R \in \mathbb{R}^{N \times N}$ (Markowitz, 1952):

$$\min_{w} \quad w^\top \Sigma_R w \tag{2.1}$$

$$\text{s.t.} \quad \mathbf{1}^\top w = 1 \tag{2.2}$$

In the absence of additional constraints, this problem exhibits a closed-form solution. However, in parts of this thesis, the portfolio optimization problem is subject to additional constraints beyond equation 2.2, for which the closed-form solution is no longer applicable. Nevertheless, as the covariance matrix is positive semi-definite, the objective function $w^\top \Sigma_R w$ is convex. Consequently, any minimum found by a local minimizer is a global minimum (Diehl, 2025, pp. 130, 138).

2.1 Covariance Matrix Estimation

In the simplest case, the covariance matrix can be estimated as the sample covariance of the returns over the previous T observations. Let $R_t \in \mathbb{R}^N$ denote the asset returns at time t. The sample covariance is then given by

$$S = \frac{1}{T-1} \sum_{t=1}^{T} (R_t - \bar{R})(R_t - \bar{R})^\top \tag{2.3}$$

However, this estimator has been shown to become unreliable when the number of assets is large relative to T, due to high estimation variance (Broadie, 1993). This may result in unstable portfolio weights, elevated transaction costs, and substantially higher realized risk than expected. Conversely, using a longer estimation window reduces variance at the expense of introducing additional bias. This stems from the non-stationary nature of asset returns, which causes the covariance matrix to change over time (Sandoval Junior & Franca, 2012). In other words, a longer estimation window produces a covariance matrix that does not accurately reflect the current covariance, but rather a smoothed historical average. This bias-variance trade-off has motivated methods to mitigate issues arising from high-dimensional estimation.

2.1.1 Ledoit-Wolf Shrinkage

An improvement over the traditional sample covariance matrix is to reduce estimation variance by deliberately introducing additional bias through shrinkage. This can be achieved by shrinking the covariance matrix towards a target matrix. One of the most widely used approaches is the Ledoit-Wolf (LW) estimator proposed by Ledoit and Wolf (2004) . It computes the estimated covariance matrix as

$$\hat{\Sigma}_{LW} = \delta F + (1 - \delta) S \tag{2.4}$$

where S denotes the sample covariance matrix, F is the target matrix and $\delta \in [0, 1]$ is the shrinkage intensity parameter. LW selects the shrinkage target as the constant-variance target:

$$F = \bar{\sigma}^2 I \tag{2.5}$$

where $\bar{\sigma}^2$ denotes the average sample variance and I is the identity matrix. The shrinkage intensity parameter δ is selected to approximately minimize the Frobenius norm loss between the estimator and the true covariance matrix. Ledoit and Wolf derive a closed-form estimator for the asymptotically optimal shrinkage coefficient:

$$\delta^* = \frac{\pi - \rho}{\gamma} \tag{2.6}$$

where π, ρ, and γ are defined as

$$\pi = \sum_{i=1}^{N} \sum_{j=1}^{N} \text{Var}\left(s_{ij}\right) \tag{2.7}$$

$$\rho = \sum_{i=1}^{N} \sum_{j=1}^{N} \text{Cov}\left(f_{ij}, s_{ij}\right) \tag{2.8}$$

$$\gamma = \|F - S\|_F^2 \tag{2.9}$$

In other words, π measures the total estimation variance of the sample covariance estimator, ρ captures the covariance between the shrinkage target and the sample covariance estimator, and γ measures the squared Frobenius distance between the sample covariance matrix and the shrinkage target.

Due to its simple implementation and strong empirical performance in high-dimensional settings, LW is a strong benchmark for covariance estimation and portfolio optimization. Furthermore, the idea of shrinking a covariance matrix towards a target is not only applicable to sample covariance matrices but also more broadly to covariance estimates characterized by low bias and high variance. However, the shrinkage intensity in equation 2.6 is specifically derived for shrinking a sample covariance estimation. In the more general case, particularly in the context of machine learning, the intensity can instead be treated as a parameter.

2.1.2 Factor Models

As an alternative approach for covariance estimation, the factor model introduced by Ross (1976) can be considered as a natural starting point. The factor model assumes that most variation in asset returns can be explained by a small number of shared factors.

In the model, the returns for asset i at time t take the general form:

$$R_t = \alpha + BF_t + u_t \tag{2.10}$$

where $\alpha \in \mathbb{R}^{N\times 1}$ is the intercept, $F_t \in \mathbb{R}^{K\times 1}$ is the zero-mean factor vector, $B \in \mathbb{R}^{N\times K}$ contains the factor loadings and $u_t \in \mathbb{R}^{N\times 1}$ is the zero-mean error term. The covariance matrix can then be derived using the assumption that $Cov(F_t, u_t) = 0$. Under these assumptions, the covariance matrix decomposes into two components (Fan et al., 2013):

$$Cov(R_t) = Cov(\alpha + BF_t + u_t) = Cov(BF_t) + Cov(u_t) = BCov(F_t)B^\top + Cov(u_t)$$

Introducing the notation $Cov(R_t) = \Sigma_R$, $Cov(F_t) = \Sigma_F$ and $Cov(u_t) = \Sigma_u$ gives a decomposition of the covariance matrix:

$$\Sigma_R = B\Sigma_F B^\top + \Sigma_u \tag{2.11}$$

Σ_F will be referred to as the systematic covariance, as it is driven by common factors, F_t, while Σ_u will be referred to as the idiosyncratic covariance, reflecting that it stems from firm-specific factors.

This is an appealing decomposition because it allows the covariance structure to be estimated in a lower-dimensional factor space, which contains significantly fewer matrix terms to estimate. Therefore, the model directly addresses the previously discussed challenges of high dimensionality.

A common approach to estimating factors F_t is to use dimensionality-reduction techniques. Such methods treat the factors that determine returns as latent, meaning they are not directly observable. They form a lower-dimensional representation of length $K << N$ to explain the returns of N assets. A simple dimensionality reduction technique is principal component analysis, PCA. PCA constructs orthogonal linear combinations, where the first principal component captures the largest possible share of the total variance of the asset's returns, the second captures the largest remaining share subject to orthogonality, and so on. The first K principal components are then selected as the latent factors for explaining the return structure (Lindholm et al., 2026, pp. 279-283). Hinton and Salakhutdinov (2006) argue that a crucial limitation of PCA and similar linear dimensionality-reduction techniques is their inability to capture non-linear rela-

tionships between the input data and the resulting latent factors. Instead, Hinton and Salakhutdinov propose using machine learning techniques, describing it as a non-linear generalization of traditional dimensionality-reduction methods. This motivates the use of neural networks for dimensionality reduction.

Chapter 3

Neural Networks for Dimensionality Reduction

This section introduces neural networks and the encoder-decoder architecture, which later serves as a basis for understanding various autoencoders and their application to covariance estimation.

3.1 Feed-Forward Neural Networks

A feed-forward neural network is a parametric mapping $f_\theta : \mathbb{R}^{H_0} \to \mathbb{R}^{H_L}$, where $H_0 = N$ denotes the input dimension, H_L the output dimension and θ the parameters (Goodfellow et al., 2016). A neural network is illustrated in figure 3.1 (Fryan et al., 2023).

For an input vector $x \in \mathbb{R}^{H_0}$, let $h^{(0)} := x$. Following Goodfellow et al. (2016), the neural network consists of layers, $h^{(l)} \in \mathbb{R}^{H_l}$, where each layer is computed as a function of the previous layer. Let L denote the total number of layers. For layers $l \in \{1, ..., L\}$, the network is recursively calculated as:

$$h^{(l)} = \sigma(W^{(l)} h^{(l-1)} + b^{(l)}) \tag{3.1}$$

$$\sigma(z^{(l)}) = max(0, z^{(l)}) \tag{3.2}$$

where $z^{(l)} = W^{(l)} h^{(l-1)} + b^{(l)}$. In particular, the middle section, consisting of layers $l \in \{1, ..., L-1\}$ of the network, is the so-called hidden layers. The network may have any number of hidden layers, each with a possibly different size. Finally, the last hidden

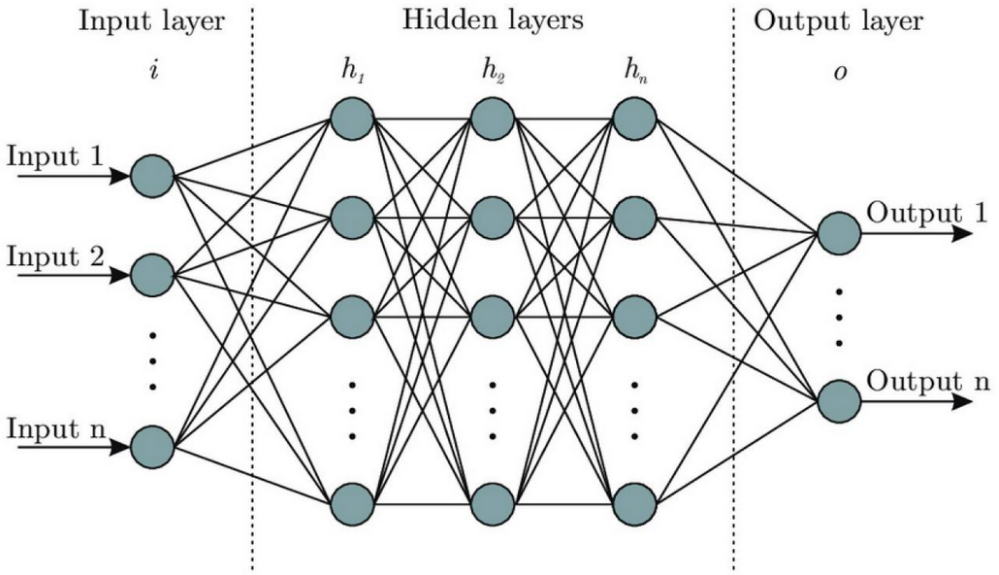

Figure 3.1: Illustration of a neural network, consisting of n inputs, three hidden layers, and n outputs.

layer is combined to create the output vector. In this setup, the trainable parameters are the set of all weights and biases, $\theta = \{W^{(l)}, b^{(l)}\}_{l=1}^{L}$.

3.1.1 Training a Neural Network

Training a neural network involves estimating the model's parameters so that it produces the desired output for a given input (Goodfellow et al., 2016). This is typically performed using a collection of input-output pairs where $y_i \in \mathbb{R}^{H_0}$ denotes the measured output, corresponding to input $x_i \in \mathbb{R}^{H_0}$. The neural network then performs the mapping:

$$\hat{y}_i := f_\theta(x_i) \tag{3.3}$$

The goal of the training process is to find parameters that minimize the discrepancy between the estimated output $\hat{y}_i$ and the true output y_i. Formally,

$$\min_\theta \frac{1}{T} \sum_{i=1}^{T} \mathcal{L}(y_i, \hat{y}_i(\theta)) \tag{3.4}$$

where $\mathcal{L}$ denotes the loss function that measures the discrepancy between the estimated and true output. To minimize it, gradient descent can be applied to the loss function, iteratively finding parameters that yield a lower loss. Let θ_i denote the parameters at

iteration i. The parameters are then updated according to

$$\theta_{k+1} = \theta_k - \eta \nabla_\theta \mathcal{L}(\theta_k) \tag{3.5}$$

where $\eta > 0$ denotes the learning rate and $\nabla_\theta \mathcal{L}(\theta_k)$ is the gradient of the loss function with respect to the parameters, evaluated at θ_k. Gradient descent can be extended to Stochastic Gradient Descent (SGD), which instead estimates the gradient using a subset of the dataset, referred to as a mini-batch, $\mathcal{B}_k \subset \{1, \ldots, T\}$. In SGD, the gradient in equation 3.5 is henceforth approximated by

$$\hat{g}_k = \frac{1}{|\mathcal{B}_k|} \sum_{i \in \mathcal{B}_k} \nabla_\theta \mathcal{L}(y_i, \hat{y}_i(\theta_k)). \tag{3.6}$$

The parameters are then updated as

$$\theta_{k+1} = \theta_k - \eta \hat{g}_k. \tag{3.7}$$

SGD can further be improved using the Adam optimizer introduced by Kingma and Ba (2014). Adam follows the same framework as SGD, but instead uses a separate learning rate η_k for each parameter. The learning rates are updated using exponentially weighted averages of past gradients and squared gradients, which estimate the first and second moments, respectively. In practice, this has been shown to yield efficient and stable convergence to the optimal parameters in high-dimensional, non-convex optimization problems.

3.1.2 Dimensionality Reduction Using Autoencoders

An autoencoder is a type of neural network that can be used to perform non-linear dimensionality reduction. Throughout the thesis, the autoencoder will be abbreviated as AE. Following Hinton and Salakhutdinov (2006) , the AE can be constructed using two sequentially connected feed-forward neural networks, called the encoder-decoder architecture. First, the encoder maps the input, $x \in \mathbb{R}^N$, into a lower-dimensional latent representation:

$$z = f_\theta(x)$$

where $z \in \mathbb{R}^K$ denotes the latent representation of the input, with $K << N$. Sub-

sequently, the decoder attempts to reconstruct the original input from the latent representation:

$$\hat{x} = g_\theta(z)$$

An AE architecture is illustrated in Figure 3.2 (Hinton and Salakhutdinov, 2006). In the illustration, an image is used as input to the encoder. This is possible because each gray pixel can be represented numerically, thereby allowing the image to be represented as a vector. The figure illustrates how an encoder compresses the input into dimensions 2000, 1000, 500, and finally a 30-dimensional representation. Thereafter, a decoder attempts to reconstruct the original image from the latent representation.

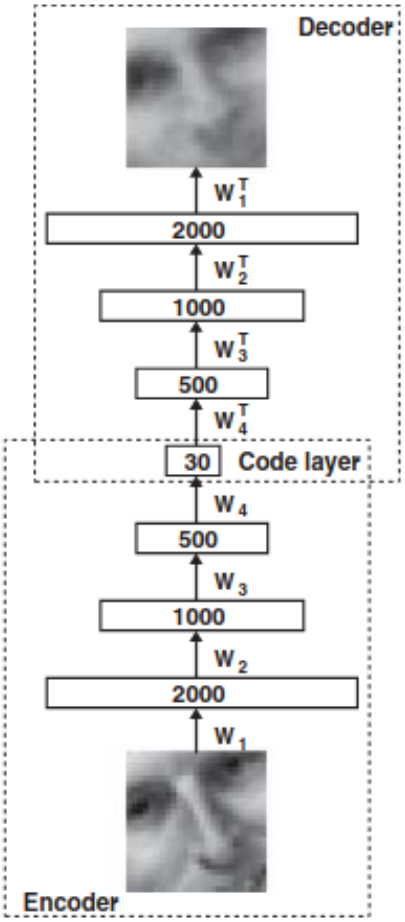

Figure 3.2: Illustration of a basic AE, consisting of an encoder and a decoder.

The autoencoder can be trained in the same manner as a traditional neural network by treating the reconstruction $\hat{x}$ as the estimated output, and x as the target, thereby minimizing reconstruction error. During training, the encoder learns to produce a latent vector that serves as a compact representation of the input since it must preserve sufficient information for the decoder to reconstruct it accurately. Since the network is only needed for dimensionality reduction, the decoder can be discarded after the encoder-decoder is trained, leaving the encoder as an optimized dimensionality-reduction mapping.

3.2 Autoencoders for Covariance Estimation

3.2.1 Static Covariance Model

The construction of covariance matrices using AEs on returns follows the approach of Conlon et al. (2025). By training an AE, it learns to compress the asset returns into a lower-dimensional space, $F_t = f_\theta(R_t) \in \mathbb{R}^K$. Thereafter, given a set of return-factor pairs, $\{(R_t, F_t)\}_{t=1}^T$, factor loadings B and intercept α are estimated using ordinary least squares:

$$R_t = \alpha + BF_t + u_t \tag{3.8}$$

where u_t is the residual error term. Notably, this construction aligns with the Ross (1976) factor model, as given in equation 2.10. Consequently, the corresponding covariance decomposition can be made, according to 2.11. However, this assumes the factors are zero-mean, $\mathbb{E}(F_t) = 0$, a condition that cannot be enforced by traditional AEs. Nevertheless, deviations from zero-mean factors only shift expected returns and therefore do not affect the covariance matrix. This becomes evident by a demean, which still yields an identical covariance decomposition to equation 2.11:

$$R_t = \alpha + BF_t + u_t \Rightarrow R_t - B\mathbb{E}(F_t) = \alpha + B(F_t - \mathbb{E}(F_t)) + u_t \tag{3.9}$$

$$\Rightarrow Cov(R_t - B\mathbb{E}(F_t)) = Cov(\alpha + B(F_t - \mathbb{E}(F_t)) + u_t) \Rightarrow \Sigma_R = B\Sigma_F B^\top + \Sigma_u \tag{3.10}$$

The systematic Σ_F and the idiosyncratic covariance Σ_u are estimated using the sample covariance matrices of the factor and residuals, respectively. Crucially, this implicitly assumes stationarity, which is an assumption without motivation that will be relaxed later. This thesis focuses on the exact factor model proposed by Fan et al. (2008)

, in which the idiosyncratic covariance matrix is constrained to be diagonal. It reflects the assumption that informative common factors will explain all systematic cross-asset dependence through Σ_F, consequently leaving only asset-specific variance in Σ_u:

$$\hat{u}_t \coloneqq R_t - \hat{\alpha} - \hat{B}F_t \tag{3.11}$$

$$\hat{\Sigma}_u = \text{diag}(\hat{\sigma}^2_{u,1}, \ldots, \hat{\sigma}^2_{u,N}), \tag{3.12}$$

Chapter 4

Proposed Conditional Autoencoder Framework

While the autoencoder forms a good basis for covariance estimation, it has fundamental limitations when it comes to adapting to different economic regimes. The autoencoder is largely agnostic to the prevailing financial environment, as it encodes asset returns without conditioning on economic information. Furthermore, while the factor model, presented in equation 2.10, assumes constant factor loadings, B, this assumption is highly restrictive in practice. It implies that each asset's sensitivity to a given latent factor is constant over time, regardless of regime changes. For instance, if the first factor corresponds to market risk and the loading is fixed at 0.05, the model dictates that the asset has 5% sensitivity to market risk, regardless of whether it is a stable period or a financial crisis. These limitations have been addressed in the asset pricing literature, where Kelly et al. (2019) propose modeling factor loadings as functions of firm characteristics. Denote $B_t \in \mathbb{R}^K$ as the factor loadings, $z_{i,t} \in \mathbb{R}^L$ as the firm characteristics of firm i at time t, and $\Gamma \in \mathbb{R}^{L \times K}$, then

$$B_t(z_{i,t-1})^\top = z_{i,t-1}^\top \Gamma \tag{4.1}$$

The approach of conditioning factor loadings on firm characteristics has proven successful in covariance estimation. See, for instance, Fan et al. (2024) and Jung (2019). However, the model by Kelly et al. (2019) is purely linear. Gu et al. (2021) argue that "There are, nonetheless, no obvious theoretical or intuitive justifications for this convenient linearity assumption. To the contrary, there are many reasons to expect that this assumption is violated." This concern is supported by the fact that essentially all leading asset pricing models predict non-linear dependencies between factors and the un-

derlying variables. Prominent examples include Campbell and Cochrane (1999), Bansal and Yaron (2004), and He and Krishnamurthy (2013).

The thesis proposes applying the conditional autoencoder to construct a nonlinear framework that conditions factor loadings on firm characteristics. The following section introduces the conditional autoencoder and thereafter constructs the framework for estimating the covariance matrix. Throughout the thesis, the conditional autoencoder will be abbreviated as CAE.

4.1 Conditional Autoencoders

The encoder side of a CAE comprises two subnetworks, each implemented as a neural network. The full architecture is shown in figure 4.1 (Gu et al., 2021))

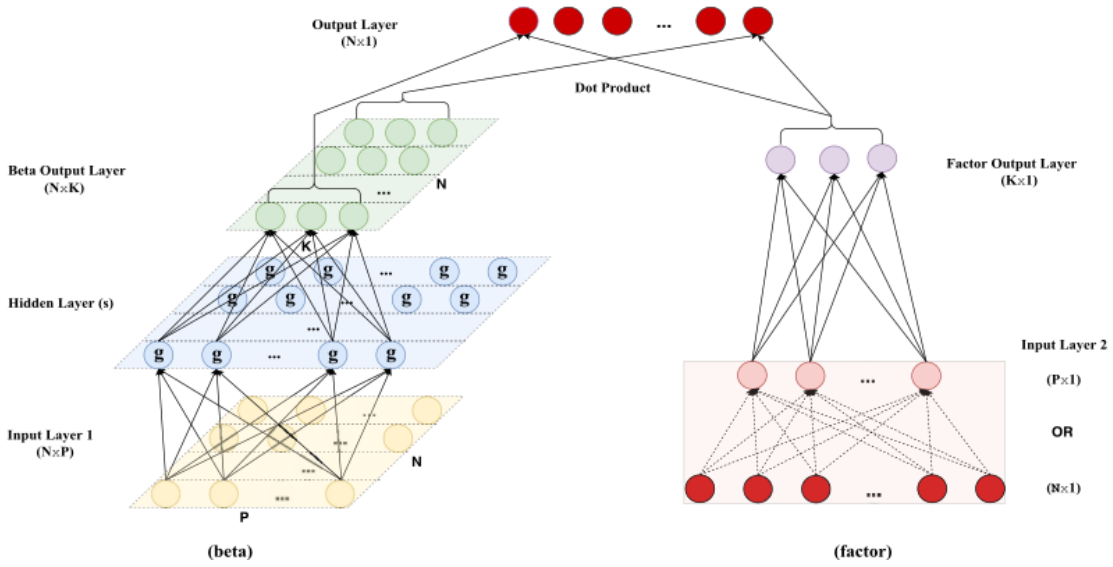

Figure 4.1: A CAE model. The figure presents a diagram of an AE incorporating firm characteristics into factor loadings and returns in factors. The right side shows factors (purple) at time t, constructed as weighted combinations of either characteristic-managed portfolios (pink) or individual asset returns (red), depending on the specification. The left side demonstrates how factor loadings (green) depend on firm characteristics (yellow) through a hidden layer (blue) with activation function g.

Let θ denote the parameters of the CAE. The neural network on the left corresponds to the beta network $\beta_\theta : \mathbb{R}^P \to \mathbb{R}^K$, which maps firm characteristics $z_{i,t-1}$ into factor

loadings. The neural network on the right corresponds to the factor component of the architecture, denoted $f_\theta : \mathbb{R}^N \to \mathbb{R}^K$. It takes N assets and outputs a latent factor of length K. Alternatively, if the number of assets and firm characteristics is high, Gu et al. (2021) propose first projecting asset returns onto the firm-characteristic space:

$$X_t = (Z_{t-1}^\top Z_{t-1})^{-1} Z_{t-1}^\top R_t \tag{4.2}$$

where $R_t \in \mathbb{R}^N$ denotes the cross-section of asset returns at time t, and $Z_{t-1} \in \mathbb{R}^{N \times P}$ is the matrix of lagged firm characteristics, where each row corresponds to a firm and each column to one of the P characteristics. X_t is thereafter used as the input to the factor network instead of individual asset returns. This substantially increases the scalability of the CAE architecture, compared to both AE and the alternative CAE method. In the study by Gu et al. (2021), roughly 30,000 stocks were projected onto 94 firm characteristics. Consequently, the dimensionality of the problem was reduced from 30,000 to 94, a stark improvement in the required network size. Moreover, such a projection lays the basis for a cleaner economic interpretation of the latent factors, which will be discussed further in the subsection "CAE interpretation".

4.1.1 Training Conditional Autoencoders

The combination of the beta and factor network corresponds only to the encoder component of the architecture. As in a standard encoder-decoder framework, the encoder must be complemented by a decoder capable of reconstructing the original returns, thereby providing a training signal for the network parameters. In a traditional AE, the decoder is typically implemented as a mirrored version of the encoder. However, this approach is not directly applicable, as the beta and factor sub-networks are jointly trained to ensure that the generated factor loadings and latent factors interact coherently to accurately reconstruct returns.

Instead of introducing a separate decoder network for each encoder, reconstruction is performed jointly. Firstly, factor loadings and factors are calculated as

$$B_t := \begin{bmatrix} \beta_\theta(z_{1,t-1})^\top \\ \beta_\theta(z_{2,t-1})^\top \\ \vdots \\ \beta_\theta(z_{N,t-1})^\top \end{bmatrix}; \quad F_t := f_\theta(R_t) \tag{4.3}$$

Then, reconstruction is performed as $B_t F_t$. Although this differs from the traditional AE reconstruction procedure, it is strongly motivated by the structure of returns according to factor models. From equation 2.10, it is evident that $B_t F_t$ constructs the systematic component of the returns. Using this reconstruction method encourages the network to encode as much information as possible within the latent factorization, so that $B_t F_t$ can closely approximate the returns. Using this architecture, the CAE can be trained using MSE between returns and $B_t F_t$ as the loss, i.e:

$$\min_{\theta} \frac{1}{T} \sum_{t=1}^{T} \|R_t - B_t F_t\|_2^2 = \min_{\theta} \frac{1}{T} \sum_{t=1}^{T} \|R_t - B_\theta(Z_{t-1}) f_\theta(R_t)\|_2^2 \tag{4.4}$$

4.1.2 Covariance Estimation from Conditional Autoencoders

Due to the structural similarities between the CAE and a traditional AE architecture, the CAE-based covariance estimator follows the general approach of an AE, with only minor modifications. Given the factor model in equation 2.10, instead of using OLS, the idiosyncratic component is rather estimated as

$$u_t = R_t - B_t(Z_{t-1}) F_t(R_t) \tag{4.5}$$

The covariance matrix is then estimated using an identical decomposition as in the traditional AE, i.e., following the theory of equation 2.11.

4.2 Dynamic Extensions

While the use of factors and factor loadings conditioned on firm characteristics provides a solid foundation for covariance estimation, some extensions can be considered. Currently, the estimated covariance matrix refers to trading month t, whereas the portfolio is held during month $t + 1$. As will be shown, these extensions also impose desirable structure on the covariance, which may further improve performance.

4.2.1 Dynamic Idiosyncratic Volatility

In the CAE framework, Σ_u represents the idiosyncratic covariance matrix, constructed from the residuals not explained by the systematic component. Since it is assumed to be diagonal, it only consists of N idiosyncratic volatilities. Each individual time series can thus be forecasted using a volatility model, such as a univariate GARCH (Bollerslev,

1986). GARCH constructs one-period forecast based on the variance persistence and past shocks as

$$\sigma^2_{u,i,t+1} = \omega_i + \alpha_i u^2_{i,t} + \beta_i \sigma^2_{u,i,t}, \quad i = 1 \dots N \tag{4.6}$$

where $\omega_i > 0$, $\alpha_i \geq 0$, and $\beta_i \geq 0$ are asset-specific coefficients that are estimated from the residuals via maximum likelihood estimation. Furthermore, returns are known to exhibit volatility clustering, meaning that periods of high volatility tend to be followed by further high volatility, and vice versa (Bollerslev, 1986). By capturing this structure, GARCH functions as a volatility filter on the noisy squared residuals.

4.2.2 Dynamic Systematic Covariance

In the CAE framework thus far, Σ_F represents the estimated covariance matrix of the systematic factors. However, in a portfolio construction scenario, a one-period forecast is more interesting. It can, for instance, be implemented using the Dynamic Conditional Correlation (DCC) model proposed by Engle (2002).

DCC is a model where individual volatilities evolve according to a GARCH model, while correlations are modeled through a separate time-varying process. The two components are subsequently combined to estimate a dynamic factor covariance matrix. The DCC framework assumes factors have a mean of zero, which the CAE does not enforce. However, solving for the factor covariance matrix of the demeaned factors will provide an equivalent solution since $Cov(F_t - \vec{\mu}) = Cov(F_t)$. DCC decomposes the covariance matrix as

$$\Sigma_{f,t+1} = D_{t+1} \Gamma_{t+1} D_{t+1} \tag{4.7}$$

where $D_{t+1} = diag(\sqrt{h_{1,t+1}}, \dots, \sqrt{h_{K,t+1}})$ contains the factor volatilities and Γ_{t+1} is a time-varying matrix of correlations between factors.

Similarly to the dynamic idiosyncratic risk model, D_{t+1} can be found by individually modeling the volatility of each factor using a univariate GARCH:

$$h_{k,t+1} = \omega_k + \alpha_k F^2_{k,t} + \beta_k h_{k,t}, \quad k = 1, \dots, K \tag{4.8}$$

where $\omega_k > 0$, $\alpha_k \geq 0$, and $\beta_k \geq 0$ are factor-specific coefficients that are estimated through maximum-likelihood estimation on each factor series.

To find Γ_{t+1}, each factor is first standardized by its estimated conditional volatility. This removes the time-varying dynamics of the factors, allowing the model to capture only the evolution of correlation between factors:

$$\epsilon_{k,t} = \frac{F_{k,t}}{\sqrt{h_{k,t}}} \tag{4.9}$$

For Γ_{t+1}, Engle (2002) introduced a method that uses standardized factors to model time-varying correlations. The model uses the same intuition as GARCH processes, where dynamics are modeled through a long-run level, recent shocks, and persistence over time. Let $a \geq 0$ denote the sensitivity to shocks, and $b \geq 0$ the persistence of correlation. The model uses a long-run target matrix $\bar{Q}$, corresponding to the sample covariance matrix over the time window. Denote the covariance matrix of the standardized factors as Q_{t+1}. Q_{t+1} is modeled as

$$Q_{t+1} = (1 - a - b)\bar{Q} + a\,\epsilon_t \epsilon_t^\top + bQ_t \tag{4.10}$$

Γ_{t+1} is then obtained from the corresponding covariance matrix Q_{t+1}:

$$\Gamma_{t+1} = diag(Q_{t+1})^{-1/2}\, Q_{t+1}\, diag(Q_{t+1})^{-1/2} \tag{4.11}$$

Lastly, the correlation matrix Γ_{t+1} and the factor volatilities are combined to form the dynamic factor-covariance matrix $\Sigma_{f,t+1}$ using equation 4.7.

4.3 Interpreting the Conditional Autoencoder Predictions

While a well-performing machine learning model can be highly useful in a portfolio management context, understanding the reasoning behind each specific prediction remains important. Interpretability can help the user assess whether a prediction is reasonable by providing insight into the underlying drivers of the prediction and enabling further evaluation of whether those drivers are reasonable. This section, therefore, presents the theory for how the output of the proposed model can be interpreted.

4.3.1 Characteristic Portfolio Attribution

Changes in the latent factors lead to changes in the estimated covariance matrix. Consequently, understanding what drives the latent factor can also provide insight into the driver behind changes in the covariance matrix. However, in one of the variations of the CAE proposed by Gu et al. (2021), returns are first projected onto the firms-characteristic space, Z, using equation 4.2. The resulting $X \in \mathbb{R}^P$ can then be interpreted as characteristic-managed portfolio returns. For instance, a column of X may correspond to the returns of a momentum portfolio. Thereafter, each latent factor consists of a combination of several different characteristic portfolios, represented as weights in $A \in \mathbb{R}^{P \times K}$:

$$F = XA \tag{4.12}$$

For instance, factor 1 may consist of 20% of a value, 30% quality, and 50% momentum exposure. This is useful for interpretability, as the model's predictions can be directly traced back to estimates of the underlying characteristic-managed portfolios.

In the alternative variation of the CAE proposed by Gu et al. (2021), in which latent factors are generated directly via a possibly nonlinear function of asset returns, the economic interpretation of these factors is not immediately clear. Nevertheless, equation 4.2 can still be used to create characteristic portfolios. Again, returns are first projected onto the firm-characteristic space, Z, using equation 4.2. However, as the network is non-linear, A must be approximated as

$$\arg\min_A \|F - XA\|_F^2 \tag{4.13}$$

While this no longer describes the exact mapping from firm characteristics to latent factors, for interpretive purposes, it gives a good idea of which underlying firm characteristics drive a latent factor.

4.3.2 Decomposing the Covariance Matrix

After the firm characteristics driving each factor have been identified, either exactly or approximately, they can be used to interpret the predictions produced by the machine learning model. For instance, if the model suddenly predicts a new regime in which correlations spike, decomposing the reasons for this prediction may be insightful.

Recall the covariance model decomposition in equation 2.11. Under the exact factor

model, all cross-asset dependencies are represented in Σ_F, meaning $B\Sigma_F B^\top$ contains all covariances attributable to shared factor risk, including both factor-specific volatility contributions and cross-factor interaction effects. This leaves only the diagonal matrix Σ_u, which contains the idiosyncratic volatilities for each asset. The systematic covariance matrix can be further decomposed into:

$$B\Sigma_F B^\top = \sum_{k=1}^{K}\sum_{l=1}^{K}[\Sigma_F]_{kl} b_k b_l^\top = \sum_{k=1}^{K}\sum_{l=1}^{K} C_{kl} \tag{4.14}$$

where b_k denotes column k of B. Thus, C_{kl} denotes the contribution of factors k and l to the systematic covariance matrix. More specifically, it captures how the covariance $[\Sigma_F]_{kl}$ propagates into asset space through the corresponding factor loadings. This decomposition can then be used to interpret changes in the covariance matrix estimated by the model. For instance, suppose the model predicts a sudden change in the covariance matrix, where the 2-norm increases significantly. This prediction can be traced to an overall increase in the 2-norm of the covariance components, C_{kl}. Furthermore, suppose the largest increase occurs in C_{33}, and A indicates that factor 3 primarily loads on quality and value. This suggests that the largest contributor to the change in the estimated covariance matrix is a predicted increase in the volatility of the synthetic quality-value portfolio.

Although $[\Sigma_F]_{33}$ corresponds to the variance of a single factor in factor space and is therefore diagonal, the resulting contribution in asset space generally affects both variances and covariances. Furthermore, since the decomposition is constructed directly from Σ_F and Σ_u, the interpretation remains applicable even when these matrices are modeled using DCC or GARCH. A brief empirical application of the interpretation framework to the actual experimental results is presented in Appendix A.

While the decomposition of Σ_F explains changes in both the covariance and variances, changes in the diagonal of the covariance matrix may additionally originate from changes in the idiosyncratic covariance matrix, Σ_u. Since Σ_u is diagonal under the exact factor model assumption, increases in its entries correspond directly to predicted increases in asset-specific idiosyncratic volatility. Consequently, when the model predicts elevated variance without substantial changes in the systematic covariance matrix, it may instead be attributed to increases in the estimated idiosyncratic volatilities.

4.3.3 Factor Return Attribution

While the covariance decomposition above provides a precise interpretation of the drivers of the covariance matrix estimate, in practice, it may be desirable to have a simpler explanation of what drives the model output. Consider equation 2.10, where the returns are driven by factor exposure, B, and latent factor realizations, F. For a given asset i, the returns can be further decomposed as

$$R_i = \alpha_i + \sum_{k=1}^{K} b_{ik} f_k + u_i, \tag{4.15}$$

where each term $b_{ik} f_k$ represents the contribution from latent factor k to the predicted return of asset i. This gives a simpler decomposition that can be combined with matrix A to understand which returns the model expects for the synthetic characteristic-portfolio corresponding to each factor. For instance, suppose a factor associated with value and quality increases greatly in its contribution $b_{ik} \odot f_k$. This suggests that the model attributes the asset's higher expected return to higher returns in the synthetic value-quality portfolio.

While this attribution does not directly explain changes in covariance, it provides a simple, intuitive interpretation of the underlying expectations that ultimately drive covariance estimation. However, unlike the previous decomposition, this interpretation does not reflect the effects introduced by the DCC or GARCH, as these are applied at a later stage.

4.4 Economic Metrics

Many different metrics exist for evaluating the performance of a minimum-variance portfolio. These metrics will be introduced in the following section.

4.4.1 Sharpe Ratio

The Sharpe ratio (SR), first introduced by Sharpe (1966), is one of the most common metrics for measuring portfolio performance. It relates the portfolio's excess returns to volatility, used as a proxy for portfolio risk. Let R_t^e denote the monthly excess returns, calculated as the difference between the monthly realized return and the monthly risk-free rate at time t. The risk-free rate is commonly estimated as the current 1-month

T-bill rate. The estimated monthly Sharpe ratio is then defined as

$$\hat{SR} = \frac{\bar{R}^e}{\sqrt{\hat{\sigma}_e^2}} \tag{4.16}$$

where the expected return and variance are estimated over the time-frame T that the Sharpe ratio is calculated over:

$$\bar{R}^e = \frac{1}{T}\sum_{t=1}^{T} R_t^e \tag{4.17}$$

$$\hat{\sigma}_e^2 = \frac{1}{T-1}\sum_{t=1}^{T}(R_t^e - \bar{R}^e)^2 \tag{4.18}$$

Typically, the Sharpe ratio is annualized by multiplying by a factor of $\sqrt{12}$.

4.4.2 Turnover

For each rebalancing date, the change in the portfolio is measured using the turnover. Let w_k denote the vector of portfolio weights at rebalance k. In line with DeMiguel et al. (2009), the turnover resulting from an isolated rebalance is then defined as:

$$TO_k = \|w_k - w_{k-1}\|_1 \tag{4.19}$$

Over a time period, the turnover is the average of individual turnovers, $\overline{TO}$.

4.5 Statistical Metrics

This section addresses the problem of comparing the similarity of two matrices. This is useful because it means that the covariance matrices estimated in a backtest, denoted $\hat{\Sigma} \in \mathbb{R}^{N\times N}$, can be compared to a target covariance, $\Sigma^{target} \in \mathbb{R}^{N\times N}$.

4.5.1 Mean-Based Loss

The most direct methods for evaluating estimation accuracy are the mean squared error (MSE) and the mean absolute error (Laurent et al., 2013). MSE is defined as

$$\text{MSE}_t = \frac{1}{N^2}\|\widehat{\Sigma}_t - \Sigma_t^{\text{target}}\|_F^2 \tag{4.20}$$

The mean absolute error, MAE, is defined as

$$\text{MAE}_t = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |\widehat{\Sigma}_{ij,t} - \Sigma_{ij,t}^{\text{target}}| \tag{4.21}$$

4.5.2 Asymmetric Loss

An asymmetric loss function is used to measure the tendency of covariance estimators to systematically overestimate the covariance matrix, following Laurent et al. (2013) . The asymmetric loss is defined as

$$\text{ASYM}_t = \frac{1}{b(b-1)} \text{tr}\left((\Sigma_t^{\text{target}})^b - \widehat{\Sigma}_t^b\right) - \frac{1}{b-1} \text{tr}\left[\widehat{\Sigma}_t^{b-1}\left(\widehat{\Sigma}_t - \Sigma_t^{\text{target}}\right)\right], \tag{4.22}$$

where $tr(\cdot)$ denotes the trace operator and $b > 1$ controls how asymmetrically the loss penalizes the overestimation compared to the underestimation. Following Laurent et al. (2013), $b = 3$ is used, corresponding to a mild degree of asymmetry.

4.5.3 Stein Loss

In practice, the target covariance matrix must itself be estimated, often as a sample covariance matrix from high-frequency data, and thus contains estimation error. Laurent et al. (2013) argue that Stein loss is a robust option for covariance evaluation against a noisy covariance target, unlike, for instance, MAE. Dey and Srinivasan (1985) define Stein loss as

$$\text{STEIN}(\hat{\Sigma}_t, \Sigma_t^{\text{target}}) = \text{tr}\Big(\hat{\Sigma}_t(\Sigma_t^{\text{target}})^{-1}\Big) - \log\Big|\hat{\Sigma}_t(\Sigma_t^{\text{target}})^{-1}\Big| - N. \tag{4.23}$$

where $\hat{\Sigma}_t \in \mathbb{R}^{N \times N}$

Chapter 5

Methodology and Data

5.1 Data

5.1.1 Datasets

The empirical analysis was conducted on an evolving universe comprising the 150 largest firms in the S&P 500 index. However, once a firm entered the universe, it was removed only after falling outside the 200 largest firms, at which point it was replaced by the largest firm not included in the universe. Excessive turnover in the firm near the cutoff, which repeatedly leaves and re-enters the universe, is thereby avoided. Furthermore, this approach allows firms experiencing major drawdowns or financial distress to temporarily remain in the sample, thereby capturing how each model behaves during periods of market stress.

Monthly data spanning 1991 to 2022 were collected. Due to data limitations, the universe only includes firms that still exist today. This means that large bankruptcies, such as Enron's, are omitted. Furthermore, mergers and acquisitions such as the Merrill Lynch acquisition are excluded from the universe. However, survivorship bias was introduced by these exclusions, as the asset universe omits firms that experienced severe financial distress. This limitation and its implications for model evaluation are discussed further in the Discussion section.

The extracted data included both return data and firm-specific characteristics. Return data was constructed as the monthly percentage change between the close of each respective month. The firm characteristics considered were the 209 characteristics provided in the open-source dataset and defined in the corresponding paper by Chen and Zimmermann (2022). However, only a subset of characteristics was selected, consisting of the following: Market capitalization, operating profitability, book-to-market ratio, return-

on-equity, earnings-to-price ratio, cash flow, cash holdings, annual book leverage, CAPM beta, 12-month momentum, 36-month mean reversion, realized volatility, idiosyncratic volatility, maximum monthly return, Amihud's illiquidity, dollar trading volume, percentage increase in assets, accruals, investment-to-revenue and share issuance. More specifically, idiosyncratic volatility was calculated as the standard deviation of the residuals from the 3-factor Fama-French model, and the realized volatility was based on the daily volatility over the past month. These were selected to cover a broad range of established predictors that had sufficient data quality.

5.1.2 Preprocessing

Following Fama and French (1992), a conservative assumption regarding data availability was made for firm characteristics. Annual data is lagged six months after its initial release. For instance, accounting data from April 2021 can be used to predict returns 18 months later, in October 2022. The authors emphasize that a six-month delay was in line with the data availability constraints of 1992, but is very conservative by modern standards. Similarly, quarterly data is lagged by one quarter.

Before the data was used in the model, it was cleaned. Extreme monthly returns were flagged and manually inspected to ensure that outliers reflect genuine market movements rather than data errors. Any returns that do not reflect genuine market movements were treated as missing. The returns were then preprocessed through a cross-sectional rank transformation, in line with Conlon et al. (2025). Let Rank$(\cdot)$: $\mathbb{R}^N \to \mathbb{R}^N$ denote a function that assigns an ordinal to each value in the vector, with the highest value in the vector being assigned N, and the lowest being assigned 1. The transformed returns are given by

$$\tilde{R}_t = \frac{Rank(R_t)}{N} - \frac{1}{2} \tag{5.1}$$

The firm characteristics were first preprocessed via imputation, with different methods applied to each characteristic. For characteristics such as 12-month momentum, which can be computed directly from returns, the values were recomputed. Slow-moving variables were forward-filled within the ticker up to 12 months. Subsequently, any remaining missing data, consisting mainly of quick-moving variables, were imputed using the median of the firm characteristics across all available stocks for the respective month. Finally, in accordance with Gu et al. (2021), the firm characteristics were transformed using a cross-sectional rank transformation before being used as model inputs.

5.2 Evaluation

The evaluation section assesses how the CAE framework performs relative to the benchmark and evaluates which variant of the CAE implementation achieves the strongest performance. The main focus was on evaluating the CAE framework, including all extensions, as it was found to give the strongest performance. However, since the standard CAE is extended with a DCC on Σ_F and a GARCH on Σ_u for forecasting and filtering, it was also tested whether these extensions actually improve performance.

The backtesting was performed on a rolling window. For each backtest, a rolling window was divided into a training, validation, and test set. The model was trained on a given combination of hyperparameters on the training set and subsequently evaluated based on their performance on the validation set. The model was then retrained on both the training and validation sets using the best hyperparameters. Thereafter, the covariance matrix was evaluated on the test set. Lastly, the rolling window was stepped forward by one month, and this process was repeated for the length of the available data.

Before the evaluation described in the following section, each covariance matrix was symmetrized to enforce that $Cov(A, B) = Cov(B, A)$, which was done using

$$\Sigma_R = \frac{\Sigma_R^T + \Sigma_R}{2}$$

Furthermore, the covariance matrix of the CAE and AE was shrunk using equation 2.4, with δ as a parameter.

5.2.1 Measuring Covariance Matrix Predictive Performance

The predictive accuracy of the estimated covariance matrices was assessed by comparing them to a target covariance matrix estimate for the given date. The target covariance matrix was constructed as the sample covariance matrix of daily returns over 12 months, centered symmetrically one month out-of-sample. The model's matrix was then evaluated against the target matrix, primarily using Stein loss and MSE, as they are robust metrics to noise in the target matrix (Laurent et al. 2013). In addition, MAE and asymmetric loss were used as secondary metrics. The performance was compared to two strong covariance benchmarks, namely the Ledoit-Wolf model, as implemented in scikit-learn , and a traditional AE as implemented by Conlon et al. (2025) . Notably, to ensure a fair comparison, the AE has all the same extensions as the proposed framework, including the DCC and GARCH.

5.2.2 Measuring Economic Performance

Based on the estimated covariance matrix of the CAE, a minimum-variance portfolio was constructed. During this test, a value-weighted portfolio was used as a naïve baseline, while minimum-variance portfolios constructed from Ledoit-Wolf and AE served as sophisticated benchmarks. This optimization was performed in two different environments. Firstly, the models were tested in an unconstrained setting. In this environment, no constraints beyond the budget constraint were imposed on the portfolio weights, and transaction costs were ignored. Since transaction costs were ignored, no turnover penalty was imposed during optimization. This evaluation complements the statistical metrics by examining whether the out-of-sample Sharpe reflects improvements in estimation. Secondly, the models were tested in a more realistic portfolio-management setting. In this case, the portfolio weights were constrained to lie between 0 and 5% of the total portfolio, and turnover was punished using the following as a penalty function,

$$\lambda \left\| w_k - w_{k-1} \right\|_2^2 \tag{5.2}$$

where w_k denotes the weights after the k-th rebalance and $\lambda > 0$ controls the strength of the turnover penalty.

Chapter 6

Results

This chapter evaluates the performance of the proposed CAE model in an out-of-sample rolling window setting. The primary objective is to assess the CAE model's performance in terms of both economic performance and statistical accuracy. As detailed in the methodology, several variations of CAE implementations are evaluated. These are denoted CAE-Full when both DCC and GARCH are included, and CAE-DCC and CAE-GARCH when only one is included, respectively. The base model, consisting only of a CAE, with no DCC or GARCH extension, is denoted CAE-Base. In each table, the value-weighted benchmark is referred to as VW, and as before, Ledoit-Wolf and the autoencoder are LW and AE, respectively.

6.1 Economic Performance

The results of this section reflect the quality of the estimated covariance matrices by using them to construct a minimum-variance portfolio.

6.1.1 Unrestricted Portfolio Performance

Throughout this specific part, transaction costs are ignored, and weights are unrestricted. Table 6.1 reports the out-of-sample performance of all models.

The CAE-based model outperforms all benchmarks in terms of the Sharpe ratio, with a Sharpe ratio of 1.01, which is 6% higher than AE, 31% higher than LW, and 60% higher than value-weighted. CAE produced the lowest drawdown at 28%.

Table 6.1: Out-of-sample gross risk-return performance (annualized).

Metric	VW	LW	AE	CAE-Full
Sharpe	0.63	0.77	0.95	1.01
Excess Return (%)	9.4	8.7	11.3	12.8
Volatility (%)	14.9	11.2	11.9	12.6
Max Drawdown (%)	-46.2	-34.8	-34.5	-28.0

6.1.2 Restricted Portfolio Performance

This section reports the results under the aforementioned realistic portfolio construction settings. It restricts weights and reports the net Sharpe ratio under transaction costs of 5 and 15 bps. The performance under these circumstances is reported in Table 6.2.

Table 6.2: Out-of-sample performance under transaction costs and weight restrictions.

Metric	VW	LW	AE	CAE-Full
Excess Return (%)	9.5	9.7	11.2	11.7
Volatility (%)	15.0	11.5	12.0	11.9
Sharpe (Gross)	0.63	0.84	0.93	0.98
Sharpe (5 bps)	0.62	0.84	0.89	0.95
Sharpe (15 bps)	0.61	0.83	0.85	0.89
Turnover (%)	3.4	4.1	28.7	28.2
Max Drawdown (%)	-47.1	-32.1	-34.6	-33.8

Across the models, the CAE model yielded the highest gross Sharpe at both 5 bps and 15 bps. It also had the highest net Sharpe under both 5 bps and 15 bps transaction costs, despite having higher turnover.

6.2 Statistical Performance

This section assesses the statistical predictive ability of the covariance estimators by comparing their respective matrices with an out-of-sample target matrix. Table 6.3 compares the statistical accuracy improvements of CAE and AE using Stein loss, MSE, MAE, and Asymmetric loss as evaluation metrics. CAE produced the lowest Stein loss, MSE, and MAE, while AE achieved the lowest asymmetric loss.

Table 6.3: Relative statistical improvement compared to Ledoit-Wolf.

Metric	AE	CAE-Full
Stein Loss	+23.1%	+24.6%
MSE	-14.8%	+5.1%
MAE	-12.5%	+5.6%
Asymmetric Loss	+5.9%	+3.5%

6.3 Evaluation of Dynamic Extensions

The framework builds on the use of DCC and GARCH to make Σ_F and Σ_u more responsive, respectively. Such framework extensions were evaluated explicitly by comparing the performance of CAE-Full, CAE-DCC, CAE-GARCH, and CAE-Base. This is done in the same manner as before, by having one constrained and one unconstrained part.

6.3.1 Economic Performance

In this comparison, each CAE variation is constructed by starting with the CAE-Full specification that yields the highest Sharpe, and thereafter iterating between excluding the DCC, GARCH, or both to construct the three other models. 6.4 reports performance under realistic portfolio management conditions. Across both transaction cost settings, 5 bps and 15 bps, CAE-Full maintained the highest net Sharpe ratio, despite having the highest turnover among all CAE variants. Among the other variants, the order remained largely unchanged after transaction costs.

Table 6.4: Out-of-sample net performance of CAE variants under transaction costs and weight restrictions.

Metric	CAE-Base	CAE-GARCH	CAE-DCC	CAE-Full
Excess Return (%)	9.9	10.2	11.4	11.7
Volatility (%)	11.6	11.7	11.8	11.9
Sharpe (Gross)	0.85	0.87	0.96	0.98
Sharpe (5 bps)	0.83	0.85	0.93	0.95
Sharpe (15 bps)	0.77	0.79	0.87	0.89
Turnover (%)	25.1	26.3	27.2	28.2

This is shown in table 6.5. Among the models, the CAE-Full achieved the highest Sharpe ratio, but only 1% better than the second-best model, CAE-DCC, that also omits the GARCH model for idiosyncratic volatilities.

Table 6.5: Out-of-sample gross risk-return performance (annualized).

Metric	CAE-Base	CAE-GARCH	CAE-DCC	CAE-Full
Sharpe	0.95	0.92	1.00	1.01
Volatility (%)	12.8	13.0	12.4	12.6
Excess Return (%)	12.1	12.0	12.4	12.8
Max Drawdown (%)	-34.3	-27.2	-28.0	-28.0

6.3.2 Statistical Performance

The section describes the statistical accuracy of each CAE variant. The results shown in table 6.6 report the statistical accuracy improvements relative to CAE-Base. Among all models, CAE-Full achieved the lowest MSE and asymmetric loss, while CAE-Full and CAE-DCC jointly achieved the lowest MAE. Furthermore, CAE-DCC achieved the lowest Stein loss.

Table 6.6: Relative improvement in statistical loss metrics compared to CAE-Base.

Metric	CAE-GARCH	CAE-DCC	CAE-Full
Stein loss	-5.5%	+1.0%	-4.2%
MSE	0.0%	+0.1%	+2.5%
MAE	-0.1%	+3.1%	+3.1%
Asymmetric Loss	-0.1%	+3.6%	+3.7%

Chapter 7

Discussion

In general, the proposed CAE framework showed clear performance improvements over the three benchmarks: the traditional AE, LW, and the value-weighted portfolio. In the unconstrained environment, this is reflected in a Sharpe ratio of 1.01 for CAE, which is 6%, 31%, and 60% higher than AE, LW, and the value-weighted portfolio, respectively. Relative to the value-weighted performance, the improvement was primarily driven by significantly lower volatility. This is unsurprising, as a minimum-variance portfolio is specifically constructed to minimize volatility, provided that the covariance matrix that CAE predicts is reasonable, which it appears to be in this case.

Compared to LW and the AE, the CAE achieved a significantly higher Sharpe ratio. Notably, in practice, if low volatility is the end goal, the CAE portfolio can be combined with a risk-free asset to scale the expected return to match that of the LW or AE portfolio, thereby achieving significantly lower volatility in the resulting portfolio (Tobin, 1958; Sharpe, 1966). Thus, even if low volatility is the objective, the higher-Sharpe CAE portfolio remains preferable, as it achieves the same volatility but higher returns.

Crucially, these improvements persist despite restrictions on portfolio weights and transaction costs. In particular, the net Sharpe ratio at 5 bps is 0.98 for the CAE framework, which is 7%, 13%, and 53% higher than those of AE, LW, and value-weighted, respectively. At 15 bps, CAE achieves a net Sharpe ratio of 0.89, corresponding to an improvement of 5%, 7%, and 46% relative to AE, LW, and VW, respectively. Lastly, the CAE framework achieves a higher predictive accuracy across all metrics than LW, most notably a 25% lower Stein loss. Compared to AE, it achieves 23% better MSE, 21% better MAE, as well as similar Stein loss and asymmetric loss.

7.1 Analysis of Performance

7.1.1 Estimation Accuracy

The increase in the Sharpe ratio is likely caused by several factors. Most directly, the CAE appears to produce a more accurate covariance forecast. This interpretation is supported by the lower error, reflected in lower Stein loss, MSE, and MAE. In minimum-variance portfolio construction, lower estimation error is particularly important because the optimization is highly sensitive to estimation errors. Michaud (1989) argued that so-called mean-variance portfolios tend to overweight assets with large and seemingly favorable estimation errors, a phenomenon known as "error maximization". As a special case of mean-variance portfolio optimization, minimum-variance optimization also exhibits this behavior through the covariance matrix, meaning that even modest improvements in covariance predictive accuracy can lead to substantial improvements in portfolio performance. Crucially, the statistical gains of the CAE model also translate into economically meaningful improvements, as reflected in consistently higher Sharpe ratios for the economically unconstrained minimum-variance portfolio. This is important, as improvements in statistical accuracy, while a good indication, can be driven primarily by small, uninteresting gains. On the contrary, the improved accuracy genuinely improves the portfolio's ability to diversify away uncompensated risk.

There are several reasons why the CAE might have achieved a better estimate. Firstly, the MVP depends on the inverse of the covariance matrix rather than on the covariance matrix itself. Consequently, even small estimation errors can be significantly amplified if the matrix is ill-conditioned (Michaud, 1989). Subsequently, this can lead to widely different performance of the minimum-variance portfolio, even if the estimates are similar. Nevertheless, looking at table 7.1, the increase in performance does not seem to be driven by an improvement in the stability of the inverse. This is evident from both the median and the maximum conditioning numbers being of the same order of magnitude. Most importantly, no model appears to produce ill-conditioned covariance matrices, meaning the difference in Sharpe is unlikely to be driven by one model producing covariance matrices with high error when inverted.

A core difference contributing to CAE's improved performance is its use of time-varying loadings that reflect the economic environment, whereas AE uses static factor loadings. This provides significant value because the covariance matrix is highly dependent on current economic conditions. For instance, in a crisis, the correlations increase towards one (Sandoval Junior & Franca, 2012). The CAE can detect this behavior from the firm characteristics and then encode them in the factor loadings. This is a stark

Table 7.1: Condition numbers of covariance estimates.

Estimator	Median $\kappa(\widehat{\Sigma})$	Max $\kappa(\widehat{\Sigma})$
CAE	151	673
AE	57	622
LW	271	519

improvement over AE, which has to infer the economic environment solely from returns. The dynamic factor loading also helps capture the fact that exposure to different risk factors changes over time, which CAE captures significantly better than AE.

Lastly, the performance difference may be driven by the responsiveness of the covariance estimates. LW is relatively static by design, as it shrinks the sample covariance towards a fixed shrinkage target based on historical measurements. In contrast, the optimal shrinkage level for CAE and AE seems to be substantially lower. This may be because the covariance matrix is estimated in a lower-dimensional representation, which may help reduce noise, while remaining adaptive. Furthermore, the CAE framework also incorporates regime conditioning by conditioning on firm characteristics. This likely helps make the model more reactive and precise than both LW and AE. Additionally, the DCC and GARCH extensions provide forecasting and volatility filtering. Consequently, the estimate may be better suited for out-of-sample evaluation than a purely historical model like LW. Nevertheless, looking at the Sharpe ratio reported for CAE-Base in table 6.5, where GARCH and DCC are omitted, it is evident that, even though the additions seem to improve performance, CAE still considerably outperforms LW without them.

7.1.2 Covariance Stability

In addition to producing better estimates, a framework may also outperform economically if it provides more stable covariance estimates. Such stability is desirable, as it typically leads to more stable portfolio weights and, consequently, lower transaction costs. Portfolio turnover reflects this stability and is therefore a crucial metric for analysis.

The results show a clear hierarchy in turnover among the models. The value-weighted portfolio achieves the lowest turnover. This is expected, as it only makes minor weight changes to reflect changes in market capitalization. Among the minimum-variance portfolios, LW has the lowest turnover, followed by CAE and AE with similar turnover. This is also expected, as LW deliberately shrinks the covariance matrix to reduce estimation variance at the cost of a slightly higher estimation bias. Furthermore, both AE and CAE include DCC and GARCH extensions in their estimation. While these increase respons-

iveness in the covariance estimate, they may also drive up transaction costs. However, the results of the different CAE variations suggest that the inclusion of such extensions leads to comparably small increases in turnover, implying that shrinkage is likely the primary driver of the lower turnover observed. This may be because the extensions also act as a filter, providing reduced noise and, consequently, lower variation.

Nevertheless, the somewhat higher turnover of CAE relative to LW and AE is justified if it yields a sufficiently large improvement in portfolio performance to offset the increase in transaction costs. Despite its higher turnover, CAE still achieved a significantly higher net Sharpe ratio under both 15- and 5-base-point transaction costs, compared to LW and AE. In other words, the improvement in the portfolio is sufficiently large to offset the higher turnover. This can be compared to AE, where transaction costs of 15 bps yield a net Sharpe ratio approximately equivalent to LW. Consequently, it is questionable whether the increase in turnover is justified for the AE. Crucially, in both cases, factors such as bid-ask spread and slippage must be taken into account when selecting a model. CAE appears to remain preferable under 15 bps in transaction costs, but for large asset managers trading illiquid small-cap stocks, a more stable model like LW may ultimately be more suitable.

7.1.3 Drawdown

In terms of drawdown, the CAE achieves a significantly lower maximum drawdown than the value-weighted portfolio at 34% versus 47%. Furthermore, in the weight-constrained portfolio, the CAE, AE, and LW achieve similar drawdowns of 34%, 35%, and 32%. However, it is difficult to conclude whether or not the slight differences in maximum drawdown between the models reflect a systematic difference in tail-risk management. Maximum drawdown is heavily influenced by a small number of extreme market movements rather than a full distribution of observations. In other words, a small difference in drawdown may reflect sample-specific outcomes rather than a consistent difference between models. In the unconstrained portfolio, however, CAE achieves a clearly lower drawdown at 28%, compared to 35% for both LW and AE. Again, such an outcome may still be sample-specific. However, since the difference is a whole 7%, it may genuinely reflect an improvement in tail-risk management for the CAE portfolio, which only appears when weights are sufficiently unconstrained to allow the CAE to exert a stronger influence on the allocation.

7.2 Performance of CAE Variations

This section analyzes the incremental contribution of each component of the framework to the covariance estimator's performance. The GARCH and DCC extensions are added separately to isolate the statistical and economic effects of forecasting idiosyncratic volatility and factor corrections.

7.2.1 Dynamic Conditional Correlation

In the CAE-Base specification, with no extensions, the systematic covariance matrix is not forecasted or filtered using DCC. Rather, it is calculated using the sample covariance. This leads to a misspecification, given that correlations vary over time; particularly during periods of financial stress, asset correlations often increase significantly (Sandoval Junior & Franca, 2012).

The results indicate that the DCC extension successfully addresses this limitation. Compared to not using any extension, DCC yields a modest improvement across all statistical metrics. However, in the economic metrics, DCC contributes greatly. Table 6.4 reports an increase of over 10% in the Sharpe ratio, when using DCC on the latent factors produced by the CAE. From a statistical perspective, this leads to a more accurate forecast of the covariance matrix. Economically, correlations are crucial to portfolio diversification. By forecasting correlation, the DCC can produce a portfolio that is better aligned with the out-of-sample economic environment. In particular, DCC can better capture correlation spikes during periods of market stress, when a sample covariance estimate would underestimate the degree of co-movement. If correlations are underestimated, the optimizer may incorrectly assume that assets provide stronger diversification benefits than they actually do. During periods of stress, this may lead the static model to allocate weights that appear well-diversified, but in practice are highly exposed to systematic risks as correlations increase.

Results show that the inclusion of DCC increases turnover. This is expected, since compared to static covariance estimators, a DCC is more responsive and thus produces covariance estimates that fluctuate more, introducing additional estimation variance. While this can produce a portfolio that is statistically closer to the target covariance matrix, it can also lead to higher turnover. Despite this, it still appears to be a substantial net gain under both 5 and 15 bps transaction costs, as seen in table 6.4. This is because DCC only increases turnover from 25% to 27%, while the gross Sharpe ratio increases by more than 10%, thus offsetting the increased transaction costs.

7.2.2 Idiosyncratic Volatility Modeling

In the CAE framework, Σ_u denotes the idiosyncratic covariance matrix, constructed from the residuals unexplained by systematic movements. As it is diagonal, it only consists of idiosyncratic variances.

The CAE-Base specification includes no forecasting or volatility filtering of the idiosyncratic volatility. Rather, it uses the sample variance of the residuals. However, volatility is known to exhibit volatility clustering, meaning that periods of high volatility tend to be followed by further periods of high volatility, and vice versa (Bollerslev, 1986). Consequently, assuming constant idiosyncratic volatility constitutes a misspecification of the covariance structure.

From a statistical standpoint, GARCH is expected to improve the estimation of the diagonal elements of the covariance estimate. In particular, it helps forecast the out-of-sample volatilities. Furthermore, it may reduce the underestimation of risk during periods of high volatility, which a sample variance would underestimate. Economically, since the minimum-variance portfolio is constructed directly from the covariance matrix, this may yield a better portfolio. Underestimating idiosyncratic volatility may cause the portfolio optimizer to allocate disproportionately large weights to assets that appear low-risk, despite actually having high volatility. Such a portfolio becomes unreasonably concentrated and risks experiencing unexpectedly large realized losses in a drawdown. However, the increased responsiveness will likely result in a covariance matrix that fluctuates more over time. As volatility forecasts change more drastically, this leads to larger weight allocation and, by extension, higher turnover.

Contrary to expectations, the GARCH extension appears to contribute relatively little to the proposed framework. The MSE, MAE, and asymmetric loss are nearly unchanged. While the most robust statistical metric, Stein loss, improves, the change is modest and does not seem to translate into economically meaningful gains in portfolio performance. More specifically, the full model is only about 1% better than the model that omits GARCH and uses sample covariance for idiosyncratic volatility modeling.

The result may be due to the interaction between the factor model and the GARCH extension. Before fitting a GARCH, a statistical test for volatility clustering was performed on each residual series. If the test fails, the framework falls back on estimating the idiosyncratic risk using the asset's sample variance. During the experiment, the GARCH test tended to succeed in periods of low volatility but failed more frequently in periods of elevated volatility. This may be because, during periods of crisis, the models seem capable of capturing a much larger share of the overall movement through the systematic component, as most firms experience simultaneous drawdowns. Consequently, much of

the dynamics may be captured by the systematic component, leaving mostly noise structure, which in turn triggers a fallback to sample variance. In other words, fallback occurs during periods of high volatility, when GARCH provides the greatest benefits relative to sample variance. Even when GARCH is successfully applied during volatile periods, the practical gains may be limited because systematic covariance prevails and dominates the overall estimate. Conversely, during periods of average volatility, more structure may exist in the idiosyncratic volatility, leading to the application of GARCH. However, in such environments, the fallback sample variance is already an acceptable estimate, limiting the benefit.

7.3 Survivorship Bias

A notable limitation of the empirical analysis is the presence of survivorship bias, caused by the data limitation of only using firms that still exist today. As aforementioned, the asset universe excludes firms that have gone bankrupt or been acquired, which tend to exhibit the highest level of financial distress.

Survivorship bias leads to a higher Sharpe ratio and a lower maximum drawdown measured for the proposed method. However, since all benchmarks use the same universe, the relative comparison minimizes this issue. Consequently, the limitation is rather whether any model is affected more by survivorship bias than others. In particular, if any model is capable of better handling firms with extreme financial distress, it is unfairly punished by survivorship bias. CAE is specifically designed to handle rapidly changing covariance, which firms in financial distress would induce. Furthermore, unlike the AE, it can capture the accompanying extreme factor loadings by conditioning on the firm's characteristics, which may reflect indicators of financial distress. On the contrary, LW relies on a static shrinkage target based on historical measurements, making it inherently unable to adapt to firms experiencing rapidly increasing financial distress. Therefore, survivorship bias likely punishes the proposed CAE more than the benchmarks. Nevertheless, if the financial distress is caused by unpredictable factors, such as large-scale accounting scandals, the exclusion of a failed firm would affect all firms equally. In such cases, the relative comparison between models would be less affected by survivorship bias, as an omitted firm's distress would be unforeseeable regardless.

Chapter 8

Conclusion

Estimating covariance matrices is a crucial yet challenging problem in portfolio optimization, particularly when the number of assets is large relative to the number of observations. The thesis contributes to the literature by utilizing the CAE model, originally proposed within asset pricing theory for conditional dimensionality reduction, to the problem of covariance estimation. More specifically, it creates a factor-model framework that uses CAEs to perform nonlinear dimensionality reduction conditioned on firm characteristics, thereby decomposing the covariance structure into a low-dimensional systematic component and an idiosyncratic component. It was extended with a dynamic conditional correlation model to forecast the systematic covariance matrix. Furthermore, GARCH models were used to forecast the estimation of the idiosyncratic volatility.

The objective of the framework was to produce covariance estimates that are both statistically accurate and economically useful. The results suggest that the proposed framework improves covariance estimation from both an economic and statistically predictive perspective, relative to several benchmarks. Economically, the minimum-variance portfolio constructed using covariance matrices generated by the proposed framework achieved a higher out-of-sample Sharpe ratio than the minimum-variance portfolio based on LW and a traditional AE, as well as a value-weighted portfolio. These improvements persisted in a realistic portfolio construction setting that incorporated transaction costs and portfolio constraints. Furthermore, across several statistical metrics, the covariance forecast from the CAE framework was closer to the target matrix calculated from out-of-sample higher-frequency returns than the benchmarks. The DCC further improved performance, indicating that dynamic covariance forecasting is a meaningful addition to the framework.

8.1 Future Work

Future research could extend this thesis in several meaningful directions. In particular, the CAE model supports two similar approaches of dimensionality reduction. The thesis mainly focuses on a method that directly compresses asset returns into a latent dimension, as this approach proved more suitable when using a small number of firm-characteristic dimensions. However, as discussed, there is an alternative method in which returns are first projected from asset space into firm-characteristic space, then compressed into the latent dimension. The method was highlighted for its strong scalability in the original CAE paper by Gu et al. (2021), where it was successfully applied to a universe of 30,000 assets. In other words, as the size of the asset universe increases, this method may become increasingly advantageous for covariance estimation. Therefore, studying the CAE framework in a universe of several thousand assets, such as the Russell 2000, could highlight the framework's scalability advantages compared to, for instance, LW.

Another possible extension is to examine whether the framework could be extended by conditioning the dimensionality reduction and covariance decomposition not only on firm characteristics, but also on macroeconomic variables. Variables such as interest rates, market volatility, and unemployment may help capture regime changes. Incorporating such information may improve the model's ability to adapt to changes in the covariance structure more accurately or swiftly.

Lastly, future research could explore more sophisticated approaches for modeling the dynamics of the systematic and idiosyncratic covariance matrices. The thesis was delimited to relatively simple approaches to forecasting, as the CAE architecture was the main focus. More advanced models have been proposed (Nelson, 1991). Examples include direct improvements to these methods, such as EGARCH, more flexible multivariate models, or stochastic volatility models. Such approaches may further improve the modeling of time-varying covariance or volatility.

Bibliography

[1] Ravi Bansal and Amir Yaron. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles". In: *The Journal of Finance* 59.4 (2004), pp. 1481–1509. DOI: https://doi.org/10.1111/j.1540-6261.2004.00670.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2004.00670.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2004.00670.x.

[2] Tim Bollerslev. "Generalized autoregressive conditional heteroskedasticity". In: *Journal of Econometrics* 31.3 (1986), pp. 307–327. ISSN: 0304-4076. DOI: https://doi.org/10.1016/0304-4076(86)90063-1. URL: https://www.sciencedirect.com/science/article/pii/0304407686900631.

[3] Mark Broadie. "Computing efficient frontiers using estimated parameters". In: *Annals of Operations Research* 45.1 (1993), pp. 21–58. ISSN: 1572-9338. DOI: 10.1007/BF02282040. URL: https://doi.org/10.1007/BF02282040.

[4] John Y. Campbell and John Cochrane. "Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior". In: *Journal of Political Economy* 107.2 (1999), pp. 205–251. DOI: 10.1086/250059. URL: https://ideas.repec.org/a/ucp/jpolec/v107y1999i2p205-251.html.

[5] Andrew Y. Chen and Tom Zimmermann. "Open Source Cross-Sectional Asset Pricing". In: *Critical Finance Review* 27.2 (2022), pp. 207–264.

[6] Thomas Conlon, John Cotter and Iason Kynigakis. "Asset allocation with factor-based covariance matrices". In: *European Journal of Operational Research* 325.1 (2025), pp. 189–203. ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2025.03.015. URL: https://www.sciencedirect.com/science/article/pii/S0377221725002127.

[7] Victor DeMiguel, Lorenzo Garlappi and Raman Uppal. "Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?" In: *The Review of Financial Studies* 22.5 (2009), pp. 1915–1953. ISSN: 08939454, 14657368. URL: http://www.jstor.org/stable/30226017 (visited on 17/05/2026).

[8] Dipak K. Dey and C. Srinivasan. "Estimation of a Covariance Matrix under Stein's Loss". In: *The Annals of Statistics* 13.4 (1985), pp. 1581–1591. ISSN: 00905364, 21688966. URL: http://www.jstor.org/stable/2241373 (visited on 31/05/2026).

[9] Stefan Diehl. *Nonlinear Optimization*. 2nd ed. Lund, Sweden: Studentlitteratur, 2025. ISBN: 9789144201139.

[10] Robert Engle. "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models". In: *Journal of Business Economic Statistics* 20.3 (2002), pp. 339–350. ISSN: 07350015. URL: http://www.jstor.org/stable/1392121 (visited on 12/03/2026).

[11] EUGENE F. FAMA and KENNETH R. FRENCH. "The Cross-Section of Expected Stock Returns". In: *The Journal of Finance* 47.2 (1992), pp. 427–465. DOI: https://doi.org/10.1111/j.1540-6261.1992.tb04398.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1992.tb04398.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1992.tb04398.x.

[12] Jianqing Fan, Yingying Fan and Jinchi Lv. "High dimensional covariance matrix estimation using a factor model". In: *Journal of Econometrics* 147.1 (2008). Econometric modelling in finance and risk management: An overview, pp. 186–197. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2008.09.017. URL: https://www.sciencedirect.com/science/article/pii/S0304407608001346.

[13] Qingliang Fan, Ruike Wu, Yanrong Yang and Wei Zhong. "Time-varying minimum variance portfolio". In: *Journal of Econometrics* 239.2 (2024), p. 105339. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2022.08.007. URL: https://www.sciencedirect.com/science/article/pii/S0304407622001646.

[14] Latefa Fryan, Mahasin Shomo and Malik Alazzam. "Application of Deep Learning System Technology in Identification of Women's Breast Cancer". In: *Medicina* 59 (Mar. 2023), p. 487. DOI: 10.3390/medicina59030487.

[15] Ian Goodfellow, Yoshua Bengio and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016, pp. 80–84, 149–150, 164–167, 170,

[16] Shihao Gu, Bryan Kelly and Dacheng Xiu. "Autoencoder asset pricing models". In: *Journal of Econometrics* 222.1, Part B (2021). Annals Issue: Financial Econometrics in the Age of the Digital Economy, pp. 429–450. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2020.07.009. URL: https://www.sciencedirect.com/science/article/pii/S0304407620301998.

[17] Zhiguo He and Arvind Krishnamurthy. "Intermediary Asset Pricing". In: *American Economic Review* 103.2 (2013), 732–70. DOI: 10.1257/aer.103.2.732. URL: https://www.aeaweb.org/articles?id=10.1257/aer.103.2.732.

[18] G. E. Hinton and R. R. Salakhutdinov. "Reducing the Dimensionality of Data with Neural Networks". In: *Science* 313.5786 (2006), pp. 504–507. DOI: 10.1126/science.1127647. eprint: https://www.science.org/doi/pdf/10.1126/science.1127647. URL: https://www.science.org/doi/abs/10.1126/science.1127647.

[19] Jaeheon Jung. *Estimating a Large Covariance Matrix in Time-varying Factor Models*. 2019. arXiv: 1910.11965 [econ.EM]. URL: https://arxiv.org/abs/1910.11965.

[20] Raymond Kan and Guofu Zhou. "Optimal Portfolio Choice with Parameter Uncertainty". In: *Journal of Financial and Quantitative Analysis* 42.3 (2007), 621–656. DOI: 10.1017/S0022109000004129.

[21] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG]. URL: https://arxiv.org/abs/1412.6980.

[22] Sébastien Laurent, Jeroen V.K. Rombouts and Francesco Violante. "On loss functions and ranking forecasting performances of multivariate volatility models". In: *Journal of Econometrics* 173.1 (2013), pp. 1–10. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2012.08.004. URL: https://www.sciencedirect.com/science/article/pii/S0304407612001777.

[23] Olivier Ledoit and Michael Wolf. "A well-conditioned estimator for large-dimensional covariance matrices". In: *Journal of Multivariate Analysis* 88.2 (2004), pp. 365–411. ISSN: 0047-259X. DOI: https://doi.org/10.1016/S0047-259X(03)00096-4. URL: https://www.sciencedirect.com/science/article/pii/S0047259X03000964.

[24] Andreas Lindholm, Niklas Wahlström, Fredrik Lindsten and Thomas B. Schön. *Machine Learning: A First Course for Engineers and Scientists*. Cambridge University Press, 2026.

[25] Harry Markowitz. "Portfolio Selection". In: *The Journal of Finance* 7.1 (1952), pp. 77–91. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2975974 (visited on 16/02/2026).

[26] Richard Michaud. "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" In: *Financial Analysts Journal - FINANC ANAL J* 45 (Jan. 1989), pp. 31–42. DOI: 10.2469/faj.v45.n1.31.

[27] Daniel B. Nelson. "Conditional Heteroskedasticity in Asset Returns: A New Approach". In: *Econometrica* 59.2 (1991), pp. 347–370. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2938260 (visited on 14/05/2026).

[28] Stephen A Ross. "The arbitrage theory of capital asset pricing". In: *Journal of Economic Theory* 13.3 (1976), pp. 341–360. ISSN: 0022-0531. DOI: https://doi.org/10.1016/0022-0531(76)90046-6. URL: https://www.sciencedirect.com/science/article/pii/0022053176900466.

[29] Leonidas Sandoval and Italo De Paula Franca. "Correlation of financial markets in times of crisis". In: *Physica A: Statistical Mechanics and its Applications* 391.1 (2012), pp. 187–208. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2011.07.023. URL: https://www.sciencedirect.com/science/article/pii/S037843711100570X.

[30] scikit-learn developers. *sklearn.covariance.LedoitWolf*. https://scikit-learn.org/stable/modules/generated/sklearn.covariance.LedoitWolf.html. (visited on 01/04/2026). 2025.

[31] William F. Sharpe. "Mutual Fund Performance". In: *The Journal of Business* 39.1 (1966), pp. 119–138. ISSN: 00219398, 15375374. URL: http://www.jstor.org/stable/2351741 (visited on 18/02/2026).

[32] J. Tobin. "Liquidity Preference as Behavior Towards Risk". In: *The Review of Economic Studies* 25.2 (Feb. 1958), pp. 65–86. ISSN: 0034-6527. DOI: 10.2307/2296205. eprint: https://academic.oup.com/restud/article-pdf/25/2/65/4427424/25-2-65.pdf. URL: https://doi.org/10.2307/2296205.

Appendix A

Interpretation of the Covariance Estimate During the Global Financial Crisis

This serves as a brief practical example of the interpretation framework presented in Chapter 4, which was implemented during rolling estimation to provide insights into the model's underlying reasoning. During the reported experiment, the CAE model predicted a substantial increase in the covariance estimate leading up to the Global Financial Crisis. The interpretation code suggests that this change was not driven by a single isolated component, but rather by a simultaneous increase in the volatility of several characteristic portfolios. In particular, factor 1 had the largest contribution to the systematic covariance matrix throughout the period, with its contribution increasing substantially at the onset of the Global Financial Crisis. Its Frobenius contribution remains relatively stable at around 0.11 before the crisis, then gradually increases, peaking at around 0.25 in 2009. The interpretation framework further suggests that factor 1 corresponds to a characteristic portfolio of primarily four firm characteristics related to volatility: realized volatility, idiosyncratic volatility, CAPM beta, and maximum monthly return. Consequently, a large share of the estimated change in the covariance matrix appeared to be driven by elevated volatility in a synthetic high-risk-characteristic portfolio. Thus, the decomposition provides an interpretable explanation for changes in the covariance estimate by linking them to synthetic portfolios rather than being purely a black-box predictor.