Estimating Industry Betas via Machine Learning: Promises and Pitfalls of Multi-Output Predictions

Tobias Cramer[1], Wolfgang Drobetz[2], and Tizian Otto[3]

October 23, 2025

Abstract

This study examines the predictive performance of multi-output machine learning models in estimating industry betas. Multi-output predictions improve forecast accuracy by identifying cross-sectional interdependencies between industries that single-output approaches systematically overlook. Two portfolio applications demonstrate the economic value of these improvements: constructing market-neutral anomaly strategies and optimizing minimum variance portfolios. Our results show that multi-output estimates enable more detailed modelling of systematic risk, leading to more effective hedging strategies, better risk management and greater alignment with investor preferences.

Keywords: Industry beta; machine learning; multi-output predictions

JEL classification codes: G11, G12, C45, C58, G17

[1] M.M.Warburg & CO and University of Hamburg Business School, Ferdinandstrasse 75, 20095 Hamburg, Germany.

[2] University of Hamburg Business School and Research Fellow at Cambridge Judge Business School, Moorweidenstrasse 18, 20148 Hamburg, Germany.

[3] Bloomberg, 3 Queen Victoria Street, London EC4N 4TQ, United Kingdom.

1 Introduction

Global equity markets are shaped by dynamic and interconnected industries whose risk profiles transform in response to macroeconomic conditions, technological disruption, and global supply chain shifts. To navigate this ever-evolving environment, financial decision makers must increasingly assess systematic risk not merely at the market or asset level, but with greater granularity, capturing sector- and industry-specific exposures. However, industry practice is predominantly anchored around the Capital Asset Pricing Model (CAPM), which continues to dominate cost of capital estimation and portfolio decision-making (Graham and Harvey, 2001; Graham, 2022). Central to this model is the market beta, a measure of a stock's sensitivity to broad market movements. Nevertheless, for many practical applications such as sector allocation, performance attribution, and risk budgeting, investors and analysts often require betas at a more disaggregated level, notably industry-level betas (Huang, O'Hara, and Zhong, 2021; Barardehi, Da, and Warachka, 2024; Biggerstaff, Goldie, and Kassa, 2025). These estimates allow decision-makers to assess sector-specific exposures, actively manage systematic risk in concentrated holdings, and align portfolios with macroeconomic views.

Despite its pertinence, the estimation of industry betas has received comparatively little attention, thus motivating the necessity for reliable industry beta estimates. Existing approaches remain limited in three essential ways: First, they rely on traditional benchmark estimators and are unable to capture the nonlinearities and complex interaction effects embedded in high-dimensional financial data. Recent work by Drobetz et al. (2024) demonstrates that market beta estimates can be improved by applying machine learning (ML) techniques capable of addressing such non-linearities and interaction effects. Second, the classification systems employed are rigid and fixed over

time, with each stock being assigned to a single industry. As a result, they either omit conglomerates that operate in multiple industries and only include firms concentrated in a single industry, which introduces a systematic upward bias into estimates, or fail to properly capture the true risk profile of multi-segment firms (Kaplan and Petersen, 1998). Third, even if considering betas towards multiple industries, each of them is estimated in isolation, thus overlooking cross-industry dependencies that may arise from shared economic exposures.

Our paper addresses these shortcomings. In particular, we propose a novel multi-output ML model to simultaneously forecast industry betas, recognizing that industry exposures are not independent. The model's capacity to discern multiple targets through shared hidden representations enables it to capture spillovers and correlated industry movements that are systematically overlooked by single-output models. In contrast to established prediction models, our approach accommodates nonlinear dynamics, captures time-varying cross-sectional interdependencies across industries, and adapts to the evolving correlation structure between industries, thereby offering a more flexible and realistic representation of systematic risk at the industry level. Multi-output ML is common in many disciplines (see Xu et al., 2019 for a broader overview), with well-established applications in computer vision (He et al., 2017), bioinformatics and genomics (Zhou and Troyanskaya, 2015), and natural language processing (Collobert et al., 2011). However, its application has received limited attention in the empirical asset pricing literature. In a related context, Richman and Scognamiglio (2024) demonstrate the effectiveness of deep learning models in simultaneously forecasting multiple yield curves, highlighting their ability to capture complex dependencies and improve forecasting accuracy.

The objective of this paper is to evaluate the predictive and economic performance of a multi-output framework for industry beta estimation, benchmarked against models that estimate each

industry beta in isolation. Our empirical framework considers a large panel of U.S. stocks, using a comprehensive set of eighty predictors including firm-level characteristics, macroeconomic predictors, dummy variables for industry membership, and sample estimates of industry betas. We rely on SIC-based industry classifications as defined by Fama and French (1997). Based on existing literature (Cosemans et al., 2016; Drobetz et al., 2024), this predictor set is designed to capture the heterogenous industry dynamics reflected in industry betas.

To the best of our knowledge, we are the first to apply ML models in the context of industry beta estimation. We benchmark their predictive performance with a set of established models from the family of rolling-window estimators. We find substantial improvement in forecast accuracy, as measured by the mean squared error (MSE), with ML models reducing forecast errors by half. This highlights their superior ability to capture nonlinearities and complex interactions within the set of predictor variables. By comparing the multi-output ML model, which estimates all industry sensitivities simultaneously (labelled "*sim*"), with the single-output ML model, which estimates all industry sensitivities separately (labelled "*sep*"), we identify improvements in forecast accuracy of up to 4.8% (value-weighted) and 6.3% (equally-weighted) across industries and time by considering cross-industry correlations or patterns. Using cross-sectional sorts based on industry beta estimates, we document that the predictive superiority of these estimates is not limited to extreme beta forecasts, but remains robust across the entire beta spectrum — from high- to low-beta stocks. Although we find heterogeneity in forecast accuracy by industry, the *sim* model outperforms or at least matches the performance of the *sep* model in every industry. Furthermore, the multi-output approach appears to be particularly advantageous for some industries such as consumer non-durables (with 9.7% improvement in the value-weighted forecast error), manufacturing (9.4%), telecommunications (8.3%), and consumer durables (5.3%).

Evaluating predictive performance over time, we find that the multi-output approach maintains a lower forecast error trajectory compared to its single-output counterpart throughout most of the sample, with pronounced outperformance during and following the dot-com bubble. In this time period, we observe a structural shift in the cross-industry correlation regime. The multi-output model architecture appears to capture and translate this shift into more precise beta forecasts. However, these improvements are not confined to isolated episodes: the *sim* model outperforms the *sep* model across various market regimes and is more often included in the "Model Confidence Set" (MCS), which contains the "best" model(s) with a given level of confidence (Hansen, Lunde and Nason, 2011). The inclusion rates reach 70.5% in telephone and television transmission and even 72.7% in manufacturing. These results highlight the effectiveness and flexibility of multi-output learning in modelling complex and time-varying risk structures, especially when shared variation across industries can be systematically exploited.

In addition to the statistical comparison, we evaluate the economic value of our industry beta forecasts through two canonical asset pricing applications: market-neutral anomaly portfolio construction and minimum variance portfolio (MVP) optimization. These settings reflect practical risk management concerns, where mitigating both market-wide and sector-specific exposures is critical. First, we extend the traditional market-neutral anomaly strategy to control for multiple sources of systematic risk. We find that multi-output forecasts deliver superior performance relative to single-output models in neutralizing unintended industry exposures, while effectively eliminating broad sector tilts. Second, we evaluate how increasingly granular representations of systematic risk affect MVP construction. By comparing optimization schemes that vary in how they incorporate systematic risk, we find that more granular representation of systematic risk enhances portfolio construction by yielding more stable and efficient minimum variance portfolios.

Our contribution to the literature is three-fold. First, we introduce a novel multi-output ML framework for estimating industry betas. This model captures nonlinear predictor interactions and structural co-movements across industries, information which has been neglected in existing approaches. Second, we examine when and how this approach adds value in terms of out-of-sample forecasting performance and portfolio implementation. Third, we address the modelling challenges that arise when applying a multi-output ML framework to asset pricing. This approach can also be applied to other prediction tasks and is valuable for both academic researchers and practitioners.

The remainder of the paper proceeds as follows: Section 2 reviews the related literature and highlights the research gap in modeling industry betas. Section 3 describes our data, and Section 4 outlines the empirical framework. Section 5 presents our statistical results, while Section 6 explores economic applications. Section 7 concludes.

2 Literature review

Ever since the CAPM of Sharpe (1964), Lintner (1965), and Mossin (1966) formalized market beta as the measure of systematic risk, the concept of beta has shaped both theoretical models and practical decision-making. It remains integral to estimating the cost of capital, evaluating portfolio risk, and designing factor-based investment strategies. The empirical estimation of precise and robust betas has received considerable attention in academic research. A variety of models have been proposed to address the challenges of measurement error, time variation, and cross-sectional heterogeneity, ranging from classic linear regressions to Bayesian shrinkage techniques and, more recently, machine learning algorithms.

The CAPM is a static, single-period model, implying that a stock's sensitivity to market risk remains constant over time. However, a large body of literature supports the notion of *time-varying* betas (Bollerslev, Engle, and Woolridge, 1988; Jagannathan and Wang, 1996; Ferson and Harvey,

1999). To address the time-varying nature, traditional beta estimation methods typically rely on rolling-window ordinary least squares (OLS) regressions (Black, Jensen, and Scholes, 1972; Fama and MacBeth, 1973). While robust to model misspecification, these traditional approaches face a bias–variance trade-off depending on the chosen window length and sampling frequency, and they are often highly sensitive to outliers. Several modifications have been proposed to address these issues. Hollstein, Prokopczuk, and Wese Simen (2019) advocate for weighted least squares using exponential weights, and Welch (2022) demonstrates that winsorizing returns can reduce forecast errors. A parallel stream of research improves beta estimates by incorporating cross-sectional information. Vasicek (1973) and Karolyi (1992) introduce shrinkage methods that pull noisy rolling betas toward priors, but these common priors may still fail to capture firm-specific nuances. Addressing this, Cosemans et al. (2016) suggest firm-level priors based on firm fundamentals. Subsequent work (Kim, Korajczyk, and Neuhierl, 2020; Kelly, Moskowitz, and Pruitt, 2021) confirms that characteristics like size, leverage, or turnover contain predictive information about future betas. Moreover, Becker et al. (2021) highlight the long-memory properties of beta time series, suggesting that persistence-based structures improve forecast stability.

Despite the advances of shrinkage and time-series estimators, recent literature increasingly highlights the benefits of machine learning for beta estimation. Drobetz et al. (2024) demonstrate that ML models, particularly tree-based models and neural networks, exhibit consistent superiority over traditional benchmark estimators from both statistical and economic perspectives. Their advantage stems largely from the ability to capture nonlinearities and complex interactions across a broad set of predictors, allowing them to approximate the unobservable true beta function more effectively. This functional flexibility leads to more stable and accurate forecasts, especially in periods of economic stress.

While *market* beta estimation has received substantial attention, the estimation of *industry* betas introduces distinct conceptual and empirical challenges that are equally relevant for systematic risk assessment. Industries serve as a natural aggregation level because firms within the same industry operate under similar economic and regulatory conditions, share common exposures to macroeconomic factors, and exhibit tightly correlated return dynamics (Moskowitz and Grinblatt, 1999). This makes industry betas particularly relevant for applications in equity valuation, capital budgeting, and portfolio risk management (Huang, O'Hara, and Zhong, 2021; Barardehi, Da, and Warachka, 2024; Biggerstaff, Goldie, and Kassa, 2025).

For U.S. stocks, there is a diverse set of industry classifications, such as the Global Industry Classification Standard (GICS), the North American Industry Classification System (NAICS), or the Standard Industrial Classification (SIC). Influential work by Fama and French (1997) establishes a SIC-based industry classification to form 48 industry portfolios[1] and estimate industry-specific exposures using both the CAPM and the three-factor model of Fama and French (1993). While their approach provides valuable insights into cross-sectional variation in the cost of equity for industries, the findings indicate imprecision in these estimates. They argue that variations in industry characteristics, which may change over time, lead to heterogeneous responses of industry betas to market-wide shocks. Baele and Londono (2013) present a more structured and rigorous treatment and focus on modeling the dynamics of industry betas using DCC–MIDAS, a model combining dynamic conditional correlation (DCC) with mixed data sampling (MIDAS), as well as kernel regression techniques. Their findings reveal both substantial persistence and time-variation in industry betas. Furthermore, the results indicate strong heterogeneity in how different sectors respond to business cycles. Unlike traditional rolling-window methods, their approach allows

[1] Alterations on this SIC-based industry classification are available on Kenneth French's website.

for flexible, data-driven weighting of historical returns and accommodates both high- and low-frequency components of beta dynamics.

A fundamental constraint of conventional industry beta estimation methodologies is the implicit assumption that each firm is exposed to merely a single industry. While this assumption is simple and intuitive, it may fail to capture the actual risk profile of diversified or multi-segment firms. Conglomerates, in particular, are either excluded or poorly represented in such frameworks. Recognizing the limitations of assigning firms to a single industry, researchers proposed segment-based models to capture the multidimensional nature of conglomerates' risk exposures. Fuller and Kerr (1981) and Kaplan and Peterson (1998) introduce the "full-information" industry beta approach, which utilizes segment-level data from multi-industry firms to infer the underlying industry betas. In this framework, a conglomerate's observed beta is interpreted as a weighted average of the unobservable true betas of its individual business segments, offering a more granular decomposition of systematic risk. Recent developments in commercial risk models acknowledge the limitations of traditional industry beta estimation and incorporate fractional industry memberships, allowing firms, especially conglomerates, to be partially allocated across multiple industries based on revenue or operating segment disclosures (Menchero and Lazanas, 2024).

Notwithstanding these advances, extant methodologies for the estimation of industry betas continue to be deficient in three principal aspects. First, existing approaches rely on historical estimators and are unable to capture the nonlinearities and complex interaction effects embedded in high-dimensional financial data. Second, models often only consider firms with operations concentrated in a single industry (pure-play industry analysis) due to their inability to handle conglomerates that operate in multiple industries. Kaplan and Petersen (1998) argue that this introduces a systematic upward bias into industry beta estimates, as conglomerates – typically firms

with higher market capitalization and lower betas – are excluded. Although the issue of diversified firm exposures is acknowledged in commercial applications, it remains underexplored in empirical asset pricing research, where firms are still commonly assigned to a single industry based on their primary line of business. Third, existing methods estimate industry betas in isolation, overlooking the cross-industry dependencies that may arise from shared economic exposures. This concern is echoed by Božović (2023), who demonstrates that industry portfolios face persistent pricing challenges due to time-varying and correlated exposures to systematic risk, and proposes a dynamic correlation factor with the market to better capture their joint behavior.

To fill this gap, we propose a multi-output approach that jointly estimates all industry betas. Our approach is designed to capture cross-sectional interdependencies across industries while accommodating the nonlinear dynamics and complex predictor interactions embedded in the predictor set. This is achieved by modelling industry betas simultaneously. For example, a positive demand shock in the semiconductor industry may simultaneously influence expected cash flows and risk exposures in downstream sectors such as consumer electronics or automotive, an effect that would be missed in models that estimate betas independently by industry. Our framework potentially offers a more flexible and realistic representation of systematic risk at the industry level.

3 Data

We obtain market data from the Center for Research in Security Prices (CRSP) and firm-level fundamentals from Compustat. The data is aggregated on a monthly frequency and denominated in U.S. dollars when currency-related. To avoid survivorship bias, we assume that firm-level fundamentals are available four months after fiscal year end, while market data becomes available immediately. Our sample includes all common stocks listed on the New York Stock Exchange (NYSE), the NYSE American (formerly known as American Stock Exchange (AMEX)), and the

National Association of Securities Dealers Automated Quotations (NASDAQ), with share codes 10 or 11. As a proxy for the risk-free rate, we use the three-month U.S. T-bill rate scaled to daily or monthly horizon. The value-weighted portfolio of all stocks is used as a proxy for the market portfolio, while value-weighted portfolios formed within each industry serve as proxies for the respective industry portfolios.

For our analysis, we rely on a comprehensive set of eighty predictors that expands upon the framework proposed by Cosemans et al. (2016) and Drobetz et al. (2024) to better capture the heterogenous industry dynamics reflected in industry betas. The predictor set includes four categories: (1) firm-level characteristics, (2) macroeconomic predictors, (3) industry membership, and (4) sample estimates of industry betas. The inclusion of a large, heterogeneous set of firm, macro, industry, and beta-based predictors allows our multi-output neural network architecture to flexibly model both idiosyncratic and cross-industry patterns in beta dynamics. Table A1 in the Appendix contains details of these predictors.

Firm-level characteristics can be further categorized into a broad set of twenty-one fundamental covariates based on accounting information as well as nine technical indicators. We assign industry dummies as predictors following the SIC-industry classification obtained from Kenneth French's Data Library clustering our sample into ten mutually exclusive industries: Consumer nondurables (*nodur*), consumer durables (*durbl*), manufacturing (*manuf*), energy (*enrgy*), business equipment (*hitec*), telephone and television transmission (*telcm*), wholesale and retail (*shops*), healthcare (*hlth*), utilities (*util*) and others (*other*). As outlined in Drobetz et al. (2024), we further include predictors based on sample estimates of industry betas over time to capture the time-series dynamics in firms' exposures to different industries. In particular, we construct three additional beta estimates for each of the ten industries. These estimates are derived using rolling-window

OLS regressions and reflect changes – rather than levels – in estimated betas. Focusing on the first differences of the estimated betas, rather than their levels, mitigates issues arising from differences in the unconditional distribution of betas across industries. For example, sectors such as *util* naturally have lower beta levels compared to sectors such as *hitec*, which tend to show higher beta levels, due to structural differences in risk exposure. These persistent differences in industry beta levels can lead to biased optimization and deteriorate predictive performance in sectors with wider or more skewed distributions of industry beta levels. The transformation ensures greater comparability and stationarity in the predictor set. It also allows for a heterogenous autoregressive forecast structure, in line with the long-memory properties documented by Becker et al. (2021). We compute the change in industry betas over three distinct horizons: three months and one year using daily returns (*ols_3m_d* and *ols_1y_d*, respectively), and five years using monthly returns (*ols_5y_m*). This results in a set of variables designed to capture short-, medium-, and long-term fluctuations in industry-specific exposures.

We adopt the same screening criteria as described by Cosemans et al. (2016). A firm is included in month *t* only if it has a nonnegative book value of equity (as defined in Fama and French (1992)), positive net sales, and a positive monthly dollar trading volume. Moreover, the firm must have return data available for the current and previous thirty-six months. Lastly, we require complete information on the predictor set. In case of missing values, we omit the entire firm-month observation to match the requirements for the econometric models. The resulting sample covers the period from April 1970 to December 2023, with 1,707 firm observations per month.

Following Cosemans et al. (2016), outliers in all firm-level characteristics are winsorized at the 0.5^{th} and 99.5^{th} percentiles of their monthly cross-sectional distributions. Finally, we rank all

characteristics each month across firms and linearly map the ranks to the (−1, +1) interval, as suggested by Kelly et al. (2019) and Freyberger, Neuhierl, and Weber (2020).

4 Methodology

Our empirical analysis examines the effectiveness of multi-output machine learning approaches in estimating industry betas. We investigate whether forecasting multiple industry betas simultaneously, rather than estimating each beta in isolation, yields more precise and robust out-of-sample predictions. The underlying rationale is that industry betas are inherently related through economic linkages and shared exposure to common risk factors. Multi-output models can exploit these interdependencies by learning the joint structure of the targets, capturing spillover effects and co-movements that single-output models systematically ignore. This section introduces our empirical framework as well as the machine learning methods and benchmark estimators.

4.1 Forecasting framework

Our empirical setup follows the methodology of Cosemans et al. (2016) and Hollstein and Prokopczuk (2016). We implement a rolling out-of-sample forecasting framework to model firm-specific industry betas. In particular, for every industry, we generate monthly forecasts of industry betas at the stock level using a rolling window approach. In each iteration, data available up to the end of month t is used to predict firm i's beta for each industry j over a forecast horizon spanning months $t+1$ to $t+k$: $\beta^{F}_{ij,t+k|t}$. We fix the forecast horizon at $k = 12$, reflecting a one-year forecast horizon. The window is then rolled forward by one month, i.e., data up to the end of $t+1$ are used to generate forecasts for $t+1+1$ to $t+1+k$. Repeating this procedure yields a panel of overlapping out-of-sample beta forecasts for each firm–industry pair, which are then evaluated against realized industry betas.

Realized industry betas are computed using daily return data over the respective one-year horizon as $\beta^R_{ij,t+k} = \frac{Cov^R_{ij,t+k}}{Var^R_{j,t+k}}$, where $Cov^R_{ij,t+k}$ denotes the realized covariance between stock i and industry j, and $Var^R_{j,t+k}$ is the realized variance of the return on the industry portfolio. Both moments are computed using continuously compounded daily returns. This approach is in line with Andersen et al. (2006), who demonstrate that realized beta measures based on high-frequency returns provide consistent estimates of the true integrated beta. For each industry j, we evaluate model performance using the value-weighted MSE at the end of each month t, defined as:

$$MSE^{(j)}_{t+k|t} = \sum_{i=1}^{N_t} w_{i,t}(\beta^R_{ij,t+k} - \beta^F_{ij,t+k|t})^2, \text{ with } k = 12, \quad (1)$$

where N_t is the number of stocks at the end of month t, $w_{i,t}$ is the market capitalization-based weight of stock i, and superscript (j) indicates the industry-specific MSE. Although future realized betas are estimates themselves, the literature supports their use as evaluation targets (Hansen and Lunde, 2006).

To obtain industry beta predictions, we use a machine learning framework that directly targets the one-year-ahead forecast of realized industry betas. In line with Gu, Kelly, and Xiu (2020a), we frame the problem as a supervised learning task, where realized betas serve as the dependent variable and a comprehensive set of firm characteristics, historical beta estimates, industry dummies, and macroeconomic variables are used as predictors. Our approach is explicitly designed to model cross-sectional variations in expected industry betas flexibly. This allows for the inclusion of a large number of potentially interacting and nonlinear predictors while maintaining the forecast-oriented objective throughout model training (Drobetz et al., 2024).

Our model is an adaptation of the additive prediction error model introduced in Gu, Kelly, and Xiu (2020a), applied to the context of industry-specific beta estimation:

$$\beta_{ij,t+k}^{R} = E_t(\beta_{ij,t+k}^{R}) + \varepsilon_{ij,t+k}, \quad (2)$$

where $\beta_{ij,t+k}^{R}$ is the realized beta of stock i with respect to industry j over the one-year forecast horizon starting at month $t+1$. The error term $\varepsilon_{ij,t+k}$ captures unpredictable components. The expected industry beta, $E_t\left(\beta_{ij,t+k}^{R}\right)$, is modeled as a function of the predictor vector $z_{i,t}$, which represents the *P*-dimensional set of predictors:

$$E_t\left(\beta_{ij,t+k}^{R}\right) = g^*\left(z_{i,t}\right). \quad (3)$$

The unknown function $g^*(\cdot)$ represents the true but unobserved mapping from predictors to future industry betas. In our analysis, we rely on neural networks to approximate $g^*(\cdot)$. Neural networks are trained to minimize the out-of-sample mean squared error (MSE) and offer a highly flexible, nonlinear, and parametric structure capable of capturing complex interactions and non-stationarities in the data. The network approximation, $g\left(z_{i,t}, \theta\right)$, depends on a high-dimensional parameter vector θ, which is estimated using the stochastic gradient descent approach and regularized through multiple techniques to prevent overfitting.

4.2 Machine learning models and network architecture

Our core modeling approach focuses on traditional feed-forward neural networks, which offer flexibility to capture complex relationships between inputs and future betas. Importantly, the prediction task can be structured in two distinct ways: (1) as a set of *separate* models, each predicting a single industry beta (single-output, "*sep*"), or (2) as a unified model that *simultaneously* predicts all industry betas (multi-output, "*sim*"). Figure A1 in the Appendix provides a visualization of these network architectures. We discuss the respective merits and design choices for these two architectures in this section.

Single-output networks are trained to predict just one target variable at a time, allowing the model to be highly specialized and optimized for that specific task. This setup can yield precise predictions for individual outputs. However, because each model is trained independently, it is not possible to exploit relationships between disparate targets. For instance, a model forecasting a firm's exposure to the *enrgy* sector will not incorporate potentially useful information from its exposure to the *manuf* sector, even though these exposures may co-move, as both industries can reflect fundamental input costs and production dependencies in the firm's operations. As a result, all knowledge about correlations or shared structure between outputs is ignored.

The appeal of multi-output networks, in contrast, is their ability to learn several targets together with a shared hidden representation, potentially capturing complex interactions among outputs that can only be handled by structured inference (Xu et al., 2019). This joint learning might improve overall *forecast accuracy*, as measured by a lower value-weighted MSE. Using industry betas in multi-output modeling is a promising approach when considering target variables as non-independent entities. This is predicated on the recognition that a firm's exposure to diverse industry sectors can be correlated, with certain industries exhibiting synchronized movement due to shared economic drivers or supply chain linkages. Moreover, training in conjunction with a shared representation can enhance *generalization*, as the presence of multiple targets serves as a form of regularization. This process functions as a guide, steering the model towards the identification of more universal predictors. This, in turn, serves to prevent the model from overfitting to the particularities of a single target (Caruana, 1997). However, if one target is noisy or sends conflicting signals compared to others, it can also degrade the model's overall performance. This underscores the importance of careful model design and implementation.

Both model architectures are built on the same foundational framework and share a common set of established regularization techniques in order to prevent overfitting and enhance generalization. These include *dropout* (randomly deactivating 5% of input predictors at each iteration), *batch normalization* (applied after the final hidden layer to stabilize activations), and *early stopping* (which halts training once the validation loss fails to improve for five consecutive epochs). In addition, we employ *learning rate scheduling* (Kingma and Ba, 2014), which reduces the learning rate gradually as convergence slows. We also use *ensembling*, in which five networks with different random seeds are trained per architecture, and their predictions are averaged to mitigate variance stemming from the stochastic optimization process. We further apply both lasso and ridge regularization to the weight parameters during training, promoting sparsity and weight shrinkage, respectively. All networks use the hyperbolic tangent (*Tanh*) activation function, defined as

$$\mathrm{TanH}(x) = \frac{2}{1+e^{-2x}} - 1, \tag{4}$$

because it delivers values in a centered range between -1 and +1. This choice aligns with the distribution of our target variables, i.e., first differences in industry betas (see Section 4.3), which are continuous, mean-reverting, and centered around zero. The centered output helps to maintain balanced gradients during training, which can accelerate convergence and improve training stability, especially when compared to alternatives like the Rectified Linear Unit (ReLU) activation function, which may be more intuitive when predicting the levels of beta estimates rather than their changes. Model training is performed with a batch size of 50, and we train each network for up to 100 epochs. Each network type uses either one, two, or three hidden layers. Network architecture follows the geometric pyramid rule (Masters, 1993), with decreasing numbers of neurons in deeper

layers.[2] A detailed summary of all model configurations and hyperparameter choices is provided in the Appendix Table A2.

Machine learning models offer considerable flexibility, but this comes at the risk of overfitting, particularly when the number of predictors is large or when the relationship between variables is unstable over time. To control model complexity and ensure valid out-of-sample inference, we tune relevant hyperparameters using time-series cross-validation. Following Gu, Kelly, and Xiu (2020a), we partition the data into rolling time windows preserving the temporal order of the data and comprising three subsamples: a training sample, a validation sample, and a test sample. For each year in the test period, we use nine years of data for training, one year for validation (and to tune hyperparameters), and reserve the following year for out-of-sample testing. We evaluate each parameter configuration by computing the time-series average of monthly value-weighted MSEs within the validation sample. The configuration that yields the lowest validation error is then selected and applied to the out-of-sample test year. We refit models once a year. A rolling-window framework captures the dynamic nature of financial data by allowing model estimates to continuously adjust as new information becomes available. Importantly, the test sample is never used during model training or parameter tuning, making it a clean benchmark for evaluating out-of-sample predictive performance.

4.3 Challenges of multi-output neural networks and corrective measures

While multi-output neural networks offer a conceptually elegant solution to predict multiple related targets simultaneously, they also face several practical and statistical challenges. Although

[2] Results not reported show that wider or deeper architectures reduce forecast accuracy in both our single- and multi-output approaches. This is consistent with Kelly, Moskowitz, and Pruitt (2021), who note that simpler networks with fewer layers and parameters tend to perform better in small samples, as deeper architectures pose greater optimization challenges due to nonconvex objectives and instability from vanishing or exploding gradients during backpropagation.

well established in disciplines such as computer vision, bioinformatics, data mining, and natural language processing (for a broader overview, see Xu et al. (2019)), to the best of our knowledge, we are the first to apply this solution in empirical asset pricing. Therefore, this section is devoted to an in-depth examination of the critical challenges associated with the implementation of multi-output architectures within the domain of asset pricing. It also delineates our strategies devised to effectively mitigate these challenges. Pitfalls can be (1) *target-related*, arising from structural heterogeneity and distributional imbalances across predicted outcomes, (2) *data-related*, stemming from sparsity, instability, or definitional ambiguity in certain industry groups, and (3) *model-related*, involving loss function specification and optimization dynamics that may inadvertently bias learning toward dominant sectors.

A fundamental challenge in multi-output prediction arises when the target variables exhibit heterogenous statistical properties. In particular, industry betas are not identically distributed and non-stationary over time, i.e., their statistical properties are subject to regime changes, structural breaks, or evolving market conditions. Panel A of Figure 1 highlights this issue by illustrating the cross-sectional dispersion of realized industry betas across all sample months. For each of the ten Fama-French industries, we display two boxplots: one for industry constituents (black bars) and another for non-constituents (gray bars). Industry constituents are defined as stocks that are exclusively assigned to a given industry based on our industry dummy variables (equal to one), whereas non-constituents are those for which the dummy equals zero. The panel reveals two major patterns. First, the cross-sectional distribution of realized industry betas is highly heterogeneous, with considerable variation in the interquartile ranges across industries. Second, there is a clear divergence between core constituents of an industry and those that are non-core constituents, i.e., stocks that

exhibit only peripheral exposure to the industry.[3] This structural heterogeneity introduces a problem for multi-output learning, where loss functions, such as MSE, typically aggregate errors uniformly across output nodes. This implicitly assumes that the distributions of all target variables are comparable in scale and dispersion. In practice, however, industry betas vary in both volatility and distributional shape. This potentially leads to biased optimization, where industries with higher variance or heavier tails disproportionately influence the optimization. The issue is further compounded by the fact that cross-sectional dispersion of industry betas is of a time-varying nature, exhibiting higher dispersion during periods of recession than during periods of expansion (Baele and Londono, 2013). To address this, we refrain from predicting levels of industry betas directly and instead model their first differences, i.e., their changes in betas. This transformation aligns the distributional properties across industries over time, improving the comparability of targets across output nodes. To retrieve levelized industry beta predictions, the predicted changes are applied to the previously observed realized industry beta, which serves as the initial anchor.

Multi-output networks depend on consistent, well-populated data across all output dimensions. Panel B of Figure 1 illustrates the time-series dispersion of value-weighted realized industry betas. While most industries exhibit stable beta distributions centered around one, two sectors – *utils* and *other* – show notably higher dispersion. This pattern may reflect structural data limitations. In the case of *utils*, the industry is consistently represented by a small number of firms, averaging fewer than ten per month, which limits the model's ability to learn stable relationships. The *other* category comprises a highly diverse and poorly defined set of firms without a coherent economic structure, making it difficult to interpret or model. The presence of such imbalanced

[3] This finding aligns with the conclusions of Kaplan and Petersen (1998), who contend that the exclusive consideration of pure play stocks engenders a systematic upward bias in industry beta estimates.

sample sizes across industries may contribute to model instability by overfitting noise in sparsely populated industries. To mitigate these problems, we exclude both *utils* and *other* from further analyses. This ensures that the model only learns from industries with statistically reliable and economically interpretable characteristics.[4]

A critical technical challenge in multi-output neural networks arises from the aggregation of individual losses across output nodes. By default, standard implementation approaches assign uniform weights to each output dimension, implicitly assuming that all targets are equally important and comparably distributed. In asset pricing, however, this assumption is problematic because the relative market share of industries changes substantially over time. Panel C of Figure 1 highlights this issue by illustrating the time-varying weights of different industries in terms of total market capitalization. Some sectors, particularly industries such as *hitec*, experienced high growth during the observation period and substantially increased their market share over time. In a standard multi-output setting with uniform loss weights, equal weights implicitly overweight larger industries in terms of the number of constituents and regardless of their market capitalization, simply because they contribute more observations to the loss function. This biases the learning process towards larger industries and potentially decreases model performance (Xu et al., 2019).

To mitigate this bias, we introduce a value-weighted loss function that applies the inverse of industry market capitalization weights, reducing the dominance of large industries and allowing for a more balanced representation across outputs. An inverse weighting scheme ensures that firms with larger market shares do not excessively drive the loss function during training. To formalize

[4] One possible remedy to mitigate sample imbalances would be to apply random oversampling, either in the form of naïve duplication based on observation counts for each industry or through a more refined value-weighted oversampling procedure that aligns sample proportions in terms of market-capitalization-based industries weights (see Xu et al. (2019) for an overview). Nevertheless, we opt for exclusion to preserve both statistical integrity and economic interpretability. In particular, oversampling structurally unstable (*utils*) or conceptually ambiguous categories (*other*) risks amplifying noise rather than reinforcing meaningful signal, thereby compromising model robustness.

this adjustment, we extend the value-weighted MSE in Equation (1) to an industry-weighted version that accounts for differences in industry size.[5] In particular, we evaluate model performance based on the IW-MSE at the end of month t, defined as:

$$IW\text{-}MSE_{t+k|t} = \sum_{j=1}^{J_t} \frac{1}{w_{j,t}} \sum_{i=1}^{N_t} w_{i,t} (\beta_{ij,t+k}^{R} - \beta_{ij,t+k|t}^{F})^2 \text{ , with } k = 12, \quad (5)$$

where N_t is the number of stocks, J the number of industries in the sample at the end of month t, and $w_{j,t}$ is the market capitalization-based weight of industry j.

In summary, while multi-output architectures offer scalability and the potential to exploit cross-target dependencies, they must be implemented with care. If left unaddressed, distributional differences, sample imbalances, and dynamic industry relevance can degrade predictive performance. Our methodological adaptations are designed to address the aforementioned challenges and ensure that the model accurately reflects the statistical and economic structure of the data.

4.4 Benchmark estimators

To evaluate the performance of our ML-based industry beta estimates, we introduce a set of benchmark estimators. These benchmarks serve two main purposes: (1) to quantify the incremental value of capturing nonlinear relationships in industry beta estimation to then (2) isolate the added value of modeling multiple outputs simultaneously, as opposed to treating betas independently. By comparing our results to these conventional approaches, we anchor our analysis in a well-established empirical framework and highlight the incremental value of machine learning methods, with particular emphasis on the benefits of multi-output modeling.

[5] While Equation (1) calculates the MSE for each industry separately, treating all industries as equally important regardless of size, the industry-weighted mean squared error (IW-MSE) in Equation (5) aggregates these individual errors using the inverse of each industry's market capitalization weight.

We apply a set of models from the family of rolling-window estimators. These benchmark models are a parsimonious alternative because they rely solely on historical return information. This circumvents the need for firm characteristics or macroeconomic predictors. This simplicity not only reduces the risk of model misspecification but also offers a clean and transparent baseline for evaluating more complex forecasting models. In particular, we focus on rolling-window estimators that compute time-varying industry betas for individual stocks via ordinary least squares (OLS) regressions of the form:

$$r_{i,ts} = \alpha_{i,t}^{H} + \beta_{ij,t}^{H} r_{j,ts} + \varepsilon_{i,ts}, \tag{6}$$

where $r_{i,ts}$ and $r_{j,ts}$ denote the excess return of stock i and the industry portfolio j, respectively. The subscript t indicates that the coefficients $\alpha_{i,t}^{H}$ and $\beta_{ij,t}^{H}$ are estimated at each point in time t based on a rolling window of daily or monthly excess returns. The subscript $s = 1, \dots, \tau$ indexes the return observations within the estimation window preceding the end of month t, where τ denotes the total length of the rolling window. The superscript H indicates that the intercept and slope coefficients are historical estimates derived from rolling-window regressions. Specifically, the intercept $\alpha_{i,t}^{H}$ captures the risk-adjusted excess return unexplained by the industry factor, while the slope $\beta_{ij,t}^{H}$ reflects stock i's historical exposure to industry j. The residual term $\varepsilon_{i,ts}$ captures idiosyncratic return variation not accounted for by the model.

We choose a set of benchmark models to address the limitations of conventional OLS regressions, which are well-documented and of considerable concern: (1) the assumption that betas remain constant within each rolling window, (2) the use of equal weighting for all observations, and (3) the high sensitivity of OLS estimates to outliers.

The first supposition posits that betas remain constant within each rolling window, thereby introducing a fundamental bias–variance trade-off. Shorter windows respond more expeditiously to changes in risk exposures; however, they yield noisier estimates. Longer windows produce more stable estimates, albeit at the cost of potentially smoothing over structural shifts. For this reason, two common benchmark estimators are constructed that vary in terms of their window lengths (τ) and data frequency. The first estimator is based on a five-year rolling window of monthly returns (*ols_5y_m*), which aligns with the work of Black, Jensen, and Scholes (1972) and Fama and Mac-Beth (1973). The second estimator is based on a one-year window of daily returns (*ols_1y_d*), which is consistent with the work of Andersen et al. (2006).

Second, to relax the assumption of constant weights within the rolling window inherent in the OLS regression and place greater weight on more recent observations, we follow Hollstein, Prokopczuk, and Wese Simen (2019) and add exponentially-weighted rolling regressions to our set of benchmark estimators. Using a one-year window of daily return data, we estimate weighted least squares (WLS) regressions, where the weights applied to observations decline exponentially over time, i.e., more recent data receives more importance. The rate of decay of the weights is determined by a half-life parameter, which establishes the rate at which the weights decrease. Specifically, we consider two settings: one with a shorter half-life (one-third of the window), labeled *ewma_s*, which reacts quickly to new information, and another with a longer half-life (two-thirds of the window), labeled *ewma_l*, which provides a more stable estimate. In both cases, the weights add up to one.

Third, to mitigate the sensitivity of OLS estimates to extreme return data, we adopt the slope winsorization approach proposed by Welch (2022). Specifically, we winsorize individual stock returns relative to the corresponding industry portfolio return, imposing the bound

$$r_{i,ts} \in (-2r_{jt,s} + 4r_{jt,s}) \tag{7}$$

under the assumption that betas plausibly fall within the interval $(-2, +4)$. We then compute industry betas using a one-year rolling window of these winsorized daily returns (*bsw*). This method effectively mitigates the impact of outliers, enhancing the robustness and stability of the resulting beta forecasts.

5 Statistical analysis

Building on the methodological framework introduced in Section 4, we now turn to an evaluation of the statistical performance of our ML models. The primary objective of our analysis is to assess the efficacy of multi-output models in predicting future industry betas, in comparison to conventional single-output alternatives. The analysis proceeds in two parts. First, we compare the forecast accuracy of industry beta forecasts across the models introduced in Section 4. Second, we examine the time-series behavior of forecast errors, with a particular focus on model performance during periods of elevated volatility and structural shifts in cross-industry relationships.

5.1 Predictive performance of industry beta estimates

We begin our empirical analysis by comparing the predictive accuracy of our forecasting models. Table 1 reports the MSE of industry beta forecasts, separately for benchmark estimators and ML models. The MSE is calculated as in Equation (1) and aggregated across time and industries, either using market capitalization-based (value-weighted) or equal-weighted schemes (Panel A) and compared across months to assess relative performance over time (Panel B). In line with Drobetz et al. (2024), we observe substantial improvement in forecast accuracy from ML models relative to traditional rolling-window estimators. Across both weighting schemes, ML models reduce MSEs by almost 50%, which highlights their superior ability to capture nonlinearities and

complex interactions within the set of predictors. Comparing both weighting schemes, the forecast accuracy of the value-weighted beta estimates is notably higher than that of their equal-weighted peers. This is likely due to the fact that equities with smaller market capitalization, which are given more weight under an equal-weighted scheme, are generally associated with more volatile beta estimates. Turning to our overall objective of comparing the multi-output model, *sim*, with the single-output model, *sep*, we observe considerable improvements in forecast accuracy of up to 4.8% and 6.3% (value-weighted and equal-weighted MSE, respectively) in terms of MSE for the multi-output approach.

Panel B further supports these findings by reporting the fraction of months during the test period (528 months) in which each model achieves a lower value-weighted average forecast error relative to its counterparts. ML models not only attain superior forecast accuracy on average but also demonstrate consistent outperformance over time. This observation underscores the reliability and practical relevance of our ML architectures. Most important, the *sim* model achieves lower forecast errors than the *sep* model in 63.83% of months. These improvements arise under identical data inputs, sample splits, and estimation procedures, which isolates the effect of the modeling architecture itself. We conclude that exploiting the interdependencies among targets by learning their joint distribution enhances predictive capability.

To understand how forecast accuracy varies at the industry level, Panel A of Figure 2 reports the value-weighted MSE of industry beta forecasts (defined in Equation (1)) for each industry and model. We observe heterogeneity in forecast accuracy by industry sector. For example, industries such as *durbl*, *hitec*, and *shops* exhibit consistently lower forecast errors, whereas errors are substantially larger for *nodur*. These differences manifest as structural in nature and are predominantly independent of the specific model implemented. Moreover, they underscore the notion that some

industries inherently exhibit elevated levels of volatility, thereby impeding the precise estimation of their betas. When comparing *sim* (black bars) and *sep* (gray bars), we observe lower forecast errors for the multi-output model because *sim* consistently matches or outperforms *sep*. To highlight this pattern, we report the percentage differences in average forecast errors (black filled triangles), calculated as one minus the ratio of the average MSE of *sim* to that of *sep*. Triangles above the line, i.e., a positive percentage difference, indicate predictive superiority of the *sim* model. The industries *nodur* (9.7%), *manuf* (9.4%), *telcm* (8.3%) and *durbl* (5.3%) appear to be particularly well-suited to benefit from the multi-output approach. Consequently, the simulation model appears to have the capacity to capitalize on interactions among target variables, thereby enhancing the accuracy of forecasts.

Panel B of Figure 2 extends the analysis to decile portfolios formed by sorting firms within each industry based on the industry beta estimate at the end of month t. For each decile, we compute the value-weighted MSE and then average across all industries and time periods. As before, *sim* is represented by black bars and *sep* by gray bars. Most important, the *sim* model consistently outperforms the *sep* model and exhibits higher predictive performance across all decile portfolios, i.e., from high- to low-beta stocks. Again, the percentage differences in average forecast errors (black filled triangles) are calculated as one minus the ratio of the average MSE of *sim* to that of *sep*. The *sim* model exhibits robust gains relative to the *sep* model across the entire beta spectrum, underscoring its overall dominance in terms of out-of-sample accuracy.

5.2 Time-series dynamics of forecast errors

After evaluating forecast accuracy across industries, we now shift our focus to the temporal dimension of model performance. Understanding how forecast accuracy evolves over time is crucial, especially in light of structural changes in market conditions and industry co-movement.

Therefore, we next analyze the time-series behavior of forecast errors to identify periods in which either model exhibits relative advantages.

Panel A of Figure 3 presents the difference in forecast errors over time. Specifically, in each month, we compute the difference between the value-weighted MSE of *sep* and that of *sim*, industry by industry, and then average across all industries. A positive forecast error indicates superior performance of the *sim* model (black bars), while negative forecast errors indicate a lower forecast error for the *sep* model (gray bars). We further highlight recession periods, as defined by the National Bureau of Economics Research (NBER). We observe that the *sim* model maintains a lower error trajectory throughout most of the sample months, with pronounced outperformance during and following the dot-com bubble.

To investigate what drives this performance gap, Panel B plots the average correlation of the target variables (industry betas), i.e., the average correlation each industry exhibits with all other industries over time. A distinct regime shift in cross-industry correlation structure becomes visible around 2000. In accordance with the findings of Baele and Londono (2013), a discernible fragmentation in the co-movement of industry betas during the dot-com bubble is evident. This fragmentation is characterized by a precipitous decline in correlations, suggesting that industries began to evolve more independently. This structural break persists into the mid-2000s and only gradually reverses following the global financial crisis. The multi-output model architecture appears to be effective in capturing and adapting to this evolving correlation structure. Specifically, its capacity to jointly model interdependent targets enables the *sim* model to translate these changes into more precise beta forecasts, particularly during periods of market turbulence. This advantage is amplified in a rolling estimation framework, where crisis episodes eventually become embedded in the model's training set, enhancing its capacity to generalize across varying market regimes.

Finally, panel C of Figure 3 extends the analysis by reporting the fraction of years in the out-of-sample period during which each model is included in the Model Confidence Set (MCS) of Hansen, Lunde, and Nason (2011). We conduct this evaluation at the yearly level to align with the annual frequency of model re-estimation. The MCS contains the "best" model(s) for each year and industry based on a given confidence level.[6] Following Becker et al. (2021), we apply a 10% significance threshold, corresponding to 90% model confidence sets. The full sample comprises 44 years. The black bars in the figure represent the *sim* model, while the gray bars represent the *sep* model. An MCS of 50% (denoted by the black line) suggests that no model has a significant advantage over the other. In accordance with its superior predictive performance across both cross-sectional and time-series dimensions, the *sim* model achieves higher MCS inclusion rates in most industries, reaching up to 70.5% in *telcm* and 72.7% in *manuf*. The only instance of relative underperformance is observed in the energy sector, wherein the *sim* model is present in the MCS in a mere 47.7% of the years. Interestingly, we also observe that the energy sector remains decoupled from the overall market over the entire observation period (as also observed in panel B), which suggests limited potential for exploiting additional information in the multi-output approach. However, this isolated case of underperformance is not statistically distinguishable from random chance as the MSC inclusion rate lies within the 95% confidence interval of a binomial distribution under the null hypothesis of equal model performance. These findings imply that the enhanced predictive performance of the *sim* model is not confined to isolated episodes but is persistent across varying market regimes. These patterns further substantiate the hypothesis that the aggregation (pooling)

[6] In many economic applications, no single model consistently outperforms all others, as data is often insufficiently informative to yield a definitive ranking. However, it is possible to narrow down the set of competing models to a smaller subset, the so-called Model Confidence Set (MCS), which contains the best-performing model(s) with a specified level of confidence. Introduced by Hansen, Lunde, and Nason (2011), the MCS procedure identifies the subset of models that cannot be statistically distinguished from the best model, where "best" is defined in terms of minimizing the mean squared error (MSE). When data is highly informative, the MCS contains a single superior model. In contrast, with less informative data, the MCS may include multiple models, reflecting greater uncertainty in model selection.

of information across beta targets enhances estimation precision, particularly in instances of substantial co-movement among industries. Conversely, the absence of such co-movement does not entail any discernible disadvantage.

More generally, our findings underscore the statistical value of employing multi-output architectures that jointly model industry betas. This is due to the fact that such architectures capture cross-target dependencies and improve out-of-sample predictive performance. This is in contrast to single-output models, which systematically ignore these dependencies. However, the efficacy of such joint modeling is contingent upon the presence of economically substantiated relationships among the targets, as only in such cases can the shared architecture prove advantageous and generate stable and dependable inputs for asset pricing.

6 Economic value

The statistical superiority of the *sim* model naturally leads us to consider its economic relevance. Improved beta forecasts are particularly valuable in applications that require market neutrality, such as beta-hedged anomaly portfolios or minimum variance portfolios. In this section, we undertake an examination of the hypothesis that the statistical improvements achieved by the multi-output approach translate into economically meaningful gains. In particular, we seek to ascertain whether there is improved hedging effectiveness and lower tracking errors in portfolio applications. In order to accomplish this objective, it is necessary to assess the economic value of our industry beta estimates. This assessment is achieved by embedding the estimates in two canonical asset pricing applications: (1) the construction of market-neutral anomaly portfolios, and (2) the optimization of MVPs. Both applications rely on the accuracy and stability of risk estimates, albeit in distinct ways: the first emphasizes hedging effectiveness and neutrality to systematic exposures, while the second focuses on variance minimization under realistic investment constraints.

6.1 Anomaly portfolio hedging

We begin our evaluation of economic value by analyzing the role of industry beta forecasts in constructing hedged anomaly portfolios. We extend the classical market-neutral framework to control for multiple sources of systematic risk simultaneously. Rather than focusing solely on market beta neutrality, we construct long-short portfolios that are ex-ante neutral with respect to a full set of industry beta exposures. This aligns more closely with practical risk management considerations, where investors should aim to eliminate broad industry tilts alongside market risk.

We consider three well-known return anomalies: size (*me*), book-to-market (*bm*), and illiquidity (*illiq*). Our portfolio optimization follows Hollstein, Prokopczuk, and Wese Simen (2019) and Drobetz et al. (2024). We start by sorting stocks into two portfolios based on the respective anomaly signal at the end of month t. Due to data limitations, we deviate from conventional decile sorting, as smaller portfolio sizes may hinder the optimization procedure's convergence. Constructing fewer but more robust signal-sorted portfolios ensures numerical stability throughout the sample. Specifically, we designate the top and bottom portfolio as the long (L) and short (S) legs, respectively. Within each leg, we use the out-of-sample beta forecasts from each model to optimize stock weights such that the ex-ante predicted industry beta matches the value-weighted industry beta profile, computed using market capitalization weights from the prior month. Due to the cross-sectional dispersion of industry betas, as outlined in Section 5, an optimization of industry exposures to equal one in both long and short legs is not always feasible. Therefore, the computation of portfolio weights for each model solves the following optimization problem independently for the long and short portfolios:

$$\min_{w_t} \sum_i \left(w_{i,t} - w_{i,t}^*\right)^2 \text{ s.t.} \tag{8}$$
$$w_{i,t} \geq 0$$
$$\sum_{j=1}^{J_t} w_{i,t} \beta_{i,\mathrm{j},t+1|t}^F = \bar{\beta}_{i,j,t}^{vw}.$$

In general, this optimization is designed to minimize the sum of the squared deviations from the initial market capitalization-based weights $w_{i,t}^*$. This choice results in more investable portfolios. The optimization is constrained such that all portfolio weights must be positive and the predicted industry betas $\beta_{i,\mathrm{j},t+1|t}^F$ must match their cross-sectional value-weighted average $\bar{\beta}_{i,j,t}^{vw}$, based on the prior month's market capitalization weights. The solution delivers portfolio weights that closely mimic the original capitalization structure, while enforcing neutrality with respect to systematic industry exposures. We combine the resulting long and short portfolios to form a hedged long–short anomaly portfolio (LS), designed to be ex-ante industry-beta neutral. The ex-post realized betas are computed as the weighted average of one-year-ahead realized industry betas, allowing us to assess the effectiveness of the hedging strategy.

Table 2 shows time-series averages of the ex-post realized industry betas for the hedged long-short (LS) anomaly portfolios. The associated *t*-statistics, shown in parentheses, are computed using Newey and West (1987) robust standard errors with eleven lags. The null hypothesis is that the respective industry beta for each long-short portfolio is equal to zero. We highlight *t*-statistics in bold when the null hypothesis cannot be rejected at the 5% significance level. In other words, this occurs when $|t| < 1.96$. While no model achieves complete ex-post neutrality across all industries, multi-output estimators deliver substantially better hedging performance. In particular, the *sim* model achieves industry neutrality, as measured by statistically insignificant exposure, in at least half of the industry dimensions across all tested anomalies. This suggests that the joint estimation of beta vectors facilitates effective control of systematic risk at a granular level.

In contrast, the *sep* model produces portfolios with systematic directional exposures across industries, tilted either positively or negatively. This pattern suggests that failing to consider cross-sectional correlation structures or co-movement of exposures can result in cumulative risk drift, which can compromise the effectiveness of hedging strategies.

These findings underscore the economic value of improved forecast accuracy in multi-output models. In particular, the joint estimation of a vector of industry betas enables the *sim* model to neutralize unintended exposures more effectively than the *sep* model. In asset pricing applications, where portfolio alignment with complex risk structures is critical, such as constructing market-neutral strategies or isolating specific factor premia, this capability offers a distinct advantage over univariate approaches.

6.2 Minimum variance portfolios

While the previous application focuses on hedging performance in anomaly-based portfolios, beta estimates also play a central role in constructing MVPs. These portfolios are particularly relevant for investors who prioritize risk control over return prediction, as MVP optimization does not depend on expected returns. Instead, optimal portfolio weights are determined entirely by the structure of the estimated covariance matrix, which itself is dependent on the quality of the beta estimates. Accurately estimating stocks' nuanced exposure to underlying risk factors, especially industry sensitivities, could greatly improve the precision of covariance forecasts. This is particularly advantageous in settings aimed at minimizing risk, where inaccurate estimation of systematic risk can result in suboptimal portfolio allocations and unintended factor tilts.

To evaluate the impact of increasingly granular risk modeling, we compare three versions of MVP construction: (1) *Market beta model*, which relies solely on market beta forecasts that are obtained using the neural net model specification as described in Section 4; (2) *Market + industry*

dummy model, which incorporates a firm's industry assignment based on a set of mutually exclusive industry dummy variables (which are equal to one for the firm's assigned industry, and zero otherwise), combined with industry-specific estimates of idiosyncratic risk; (3) *Market + industry beta model*, which uses the industry beta estimates for each stock derived from the multi-output approach, alongside their respective industry-specific idiosyncratic risk estimates. This model progression enables the isolation of the incremental value of modeling nuanced industry exposures when estimating the covariance matrix. The third specification, based on a full set of jointly estimated industry betas, offers the richest and most flexible representation of systematic risk.

Following Cosemans et al. (2016), we assume a single-factor structure to approximate the high-dimensional covariance matrix. At the end of each month t, we forecast the out-of-sample covariance matrix for month $t+1$ as:

$$\Omega_{t+1|t} = s^2_{M,t+1|t} B_{t+1|t} B'_{t+1|t} + D_{t+1|t}, \tag{9}$$

where $B_{t+1|t}$ is the $N_t \times 1$ vector of out-of-sample beta forecasts, $s^2_{M,t+1|t}$ is the predicted market variance (taken as the variance of daily market excess returns over the prior year), and $D_{t+1|t}$ is the diagonal matrix of predicted idiosyncratic variances for each stock. The idiosyncratic variances are computed based on daily returns over the past year ending in month t and using the residuals of a one-factor model, $r_{i,t} - \beta^F_{i,t} r_{M,t}$. These estimates are assumed to persist into the next month. Using this estimated covariance matrix, we construct the MVP by solving the following optimization problem:

$$\begin{aligned} &\min_{w_t} \; w'_t \Omega_{t+1|t} w_t \;\; \text{s.t.} \\ &0 \le w_{i,t} \le 0.05 \\ &\textstyle\sum_{i=1}^{N_t} w_{i,t} = 1. \end{aligned} \tag{10}$$

The constraints ensure that the portfolio is fully invested and individual stock weights remain within a realistic range, i.e., between 0 and 5%, consistent with short-selling restrictions and position limits typically imposed in institutional settings. The portfolio is rebalanced monthly, and we track its out-of-sample performance using realized returns in month $t + 1$. Return and risk statistics are computed based on these monthly portfolio returns.

Next, we extend the covariance matrix by incorporating industry-level risk via a set of industry dummies, defined according to the Fama and French (1997) SIC-based classification. This introduces an industry risk component into the optimization framework, effectively augmenting the market risk model in Equation (9) to a multi-factor model:

$$\Omega_{t+1|t} = S_{t+1|t} B_{t+1|t} B'_{t+1|t} + D_{t+1|t}, \tag{11}$$

where $B_{t+1|t}$ is an $N_t \times K$ matrix of factor loadings reflecting market beta and industry dummies, and $S_{t+1|t}$ is a $K \times K$ diagonal matrix of factor variances computed from the variance of daily market and industry excess returns over the previous year. Idiosyncratic variances are computed based on the residuals from the factor model: $r_{i,t} = \sum_{j=1}^{K} \delta_{ij} r_{j,t} + \varepsilon_{i,t}$, where δ_{ij} = 1 if firm i belongs to industry j, and 0 otherwise, $r_{j,t}$ is the value-weighted return of industry j, and $r_{i,t}$ is the return of firm i. Under this specification, the covariance matrix reflects both market-wide and industry-specific sources of systematic risk, with the implicit assumption that industry betas are constant and equal to one.

Finally, we generalize this framework further by replacing the binary industry dummies with our industry beta estimates derived from the multi-output approach. Unlike the simpler dummy-based specification, this approach allows each stock i to exhibit continuous and time-varying exposure to all industries. This captures a richer, more nuanced structure of systematic risk. Formally, the factor loading matrix $B_{t+1|t}$ now contains the market beta and industry beta estimates for each

stock i, and the residuals used to construct $D_{t+1|t}$ are obtained from a linear projection of stock returns onto the full set of industry factor returns: $r_{i,t} = \sum_{j=1}^{K} \beta_{i,j,t}^{F} r_{j,t} + \varepsilon_{i,t}$, where $\beta_{i,j,t}^{F}$ denotes the industry beta estimate of stock i for industry j, reflecting each asset's unique loading on industry-level factors.

Table 3 summarizes the risk characteristics of the MVPs. All performance metrics are based on monthly portfolio returns. We observe that the ex-post standard deviation (*Std*) of the MVPs decreases monotonically as the modeled risk exposures become more granular. This trend holds not only for overall volatility but also across multiple dimensions of risk. In particular, the downside deviation (*Dwnd*), which is computed over months with negative returns, the lowest monthly excess return (*Min*), and the maximum drawdown (*MaxDD*), defined as the largest cumulative loss from peak to trough, improve substantially when transitioning from market-only to multi-industry risk models. Our results confirm the hypothesis that more accurate representations of systematic risk directly enhance portfolio stability. Most important, incorporating multi-output industry betas allows for a more nuanced decomposition of common risk exposures, which in turn leads to more effective variance minimization.

Our findings demonstrate that the observed statistical improvements in industry beta estimation translate into economically meaningful gains. The joint estimation of the full vector of industry betas is a methodological advancement that enables multi-output models to systematically identify cross-industry dependencies and nonlinear risk structures that are typically overlooked by single-output models. This more granular representation of systematic risk directly enhances portfolio construction, whether by improving hedging precision in anomaly strategies or by yielding more stable and efficient minimum variance portfolios. It is evident across both applications that more

accurate and robust industry beta forecasts facilitate a more optimal alignment of portfolio exposures with investors' risk objectives.

7 Conclusion

We compare the predictive performance of multi-output neural networks for industry beta estimation against both conventional single-output neural network and benchmark estimators. Using a comprehensive panel of U.S. stocks, we find that the multi-output model not only improves out-of-sample forecast accuracy but also delivers more stable and economically relevant industry beta estimates, particularly during periods of crisis with structural shifts in cross-industry co-movements. In economic applications, the multi-output neural network model enhances hedging precision in anomaly portfolios by achieving broader industry neutrality and improves minimum variance portfolio construction through a more granular representation of systematic risk.

We conclude that by simultaneously estimating industry betas, the multi-output model can capture cross-sectional interdependencies across industries, while also accommodating the nonlinear dynamics and complex predictor interactions embedded in the predictor set. However, to fully realize these benefits, such architectures must be implemented with care, as distributional shifts, sample imbalances, or changes in industry relevance can degrade their predictive performance if left unaddressed.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Wu, J. (2006). Realized Beta: Persistence and Predictability. In T. Fomby, & D. Terrel, *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series.* Amsterdam, Netherlands: Elsevier.

Baele, L., & Londono, J. M. (2013). Understanding Industry Betas. *Journal of Empirical Finance, 22*, 30-51.

Barardehi, Y. H., Da, Z., & Warachka, M. (2024). The Information in Industry-Neutral Self-Financed Trades. *Journal of Financial and Quantitative Analysis, 59*(2), 796-829.

Becker, J., Hollstein, F., Prokopczuk, M., & Sibbertsen, P. (2021). The Memory of Beta. *Journal of Banking and Finance, 124*(1), 106026.

Biggerstaff, L., Goldie, B., & Kassa, H. (2025). Beta Estimation Precision and Corporate Investment Efficiency. *Journal of Corporate Finance, 91*(1), 102728.

Black, F., Jensen, M. C., & Scholes, M. S. (1972). The Capital Asset Pricing Model: Some Tests. In M. C. Jensen, *Studies in the Theory of Capital Markets.* New York (NY), U.S.: Praeger.

Bollerslev, T., Engle, R. F., & Wooldridg, J. M. (1988). A Capital Asset Pricing Model with Time-Varying Covariances. (131, Ed.) *Journal of Political Economy, 96*(1), 116.

Božović, M. (2023). Can a Dynamic Correlation Factor Improve the Pricing of Industry Portfolios? *Finance Research Letters, 53*, 103626.

Caruana, R. (1997). Multitask Learning. *Machine Learning, 28*, 41-75.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research, 12*(12), 2493-2537.

Cosemans, M., Frehen, R., Schotman, P. C., & Bauer, R. (2016). Estimating Market Betas Using Prior Information Based on Firm Fundamentals. *The Review of Financial Studies, 29*(4), 1072-1112.

Drobetz, W., Hollstein, F., Otto, T., & Prokopczuk, M. (2024). Estimating Stock Market Betas via Machine Learning. *Journal of Financial and Quantitative Analysis*, 1-37.

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance, 47*(2), 427-465.

Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics, 33*(1), 3-56.

Fama, E. F., & French, K. R. (1997). Industry Costs of Equity. *Journal of Financial Economics, 43*(2), 153-193.

Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy, 81*(3), pp. 607-636.

Ferson, W. E., & Harvey, C. R. (1999). Conditioning Variables and the Cross Section of Stock Returns. *The Journal of Finance, 54*(4), 1325-1360.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting Characteristics Nonparametrically. *The Review of Financial Studies, 33*(5), 2326-2377.

Fuller, R. J., & Kerr, H. S. (1981). Estimating the Divisional Cost of Capital: An Analysis of the Pure-Play Technique. *The Journal of Finance, 36*(5), 997-1009.

Graham, J. R. (2022). Presidential Address: Corporate Finance and Reality. *The Journal of Finance, 77*(4), 1975-2049.

Graham, J. R., & Harvey, C. R. (2001). The Theory and Practice of Corporate Finance: Evidence from the Field. *Journal of Financial Economics, 60*(2-3), 187-243.

Gu, S., Kelly, B., & Xiu, D. (2020a). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies, 33*(5), 2223-2273.

Hansen, P. R., & Lunde, A. (2006). Consistent Ranking of Volatility Models. *Journal of Econometrics, 131*(1-2), 97-121.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The Model Confidence Set. *Econometrica, 79*(2), 453-497.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *IEEE International Conference on Computer Vision (ICCV)*, 2980-2988.

Hollstein, F., & Prokopczuk, M. (2016). Estimating Beta. *The Journal of Financial and Quantitative Analysis, 51*(4), 1437-1466.

Hollstein, F., Prokopczuk, M., & Wese Simen, C. (2019). Estimating Beta: Forecast Adjustments and the Impact of Stock Characteristics for a Broad Cross-Section. *Journal of Financial Markets, 44*(1), 1437-1466.

Huang, S., O'Hara, M., & Zhong, Z. (2021). Innovation and Informed Trading: Evidence from Industry ETFs. *Review of Financial Studies, 34*(3), 1280-1316.

Jagannathan, R., & Wang, Z. (1996). The Conditional CAPM and the Cross-Section of Expected Returns. *The Journal of Finance, 51*(1), 3-53.

Kaplan, P. D., & Peterson, J. D. (1998). Full-Information Industry Betas. *Financial Management, 27*(2), 85-93.

Karolyi, G. A. (1992). Predicting Risk: Some New Generalizations. *Management Science, 38*(1), 57-74.

Kelly, B. T., Moskowitz, T. J., & Pruitt, S. (2021). Understanding Momentum and Reversal. *Journal of Financial Economics, 140*(3), 726-743.

Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are Covariances: A Unified Model of Risk and Return. *Journal of Financial Economics, 134*(3), 501-524.

Kim, S., Korajczyk, R. A., & Neuhierl, A. (2021). Arbitrage Portfolios. *The Review of Financial Studies, 34*(6), 2813-2856.

Kingma, D. P., & Ba, J. L. (2014). Adam: A Method for Stochastic Optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR).* Banff.

Lintner, J. (1965). Security Prices, Risk, and Maximal Gains From Diversification. *The Journal of Finance, 20*(4), 587-615.

Masters, T. (1993). *Practical Neural Network Recipes in C++*. Burlington (MA), US: Morgan Kaufmann Publishers.

Menchero, J., & Lazanas, A. (2024). Evaluating and Comparing Risk Model Performance. *The Journal of Portfolio Management Quantitative Special Issue, 50*(3), 90-110.

Moskowitz, T. J., & Grinblatt, M. (1999). Do Industries Explain Momentum? *The Journal of Finance, 54*(4), 1249-1290.

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica, 34*(4), 768-783.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica, 55*(3), 703-708.

Richman, R., & Scognamiglio, S. (2024). Multiple Yield Curve Modeling and Forecasting Using Deep Learning. *Astin Bulletin, 54*, 463-494.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance, 19*(3), 425-442.

Vasicek, O. A. (1973). A Note on Using Cross-Sectional Information in Bayesian Estimation of Market Betas. *The Journal of Finance, 28*(5), 1233-1239.

Welch, I. (2022). Simply Better Market Betas. *Critical Finance Review, 11*(1), 37-64.

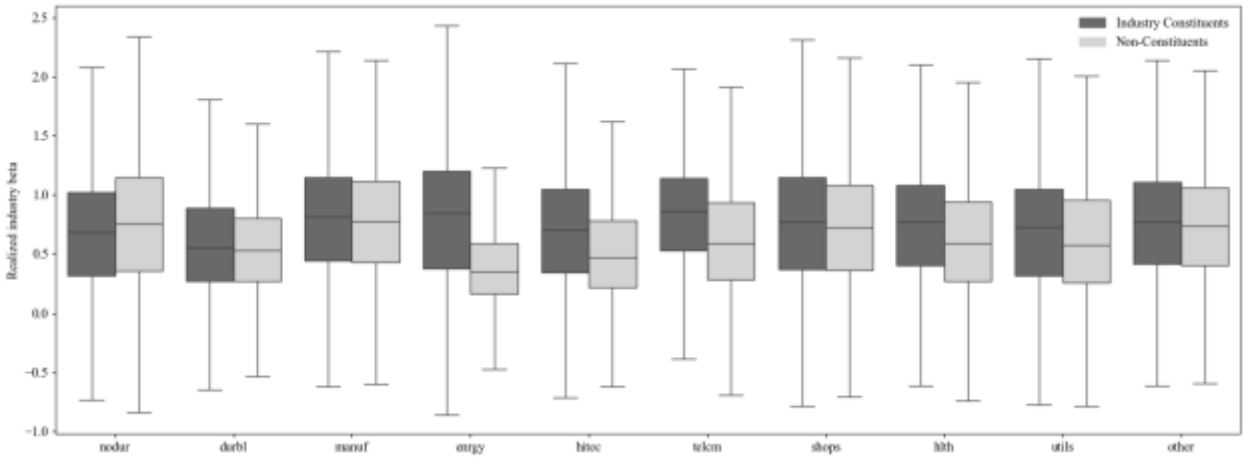
Xu, D., Shi, Y., Tsang, I. W., Ong, Y.-S., Gong, C., & Shen, X. (2019). Survey on Multi-Output Learning. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(7), 2409-2429.

Zhou, J., & Troyanskaya, O. G. (2015). Predicting Effects of Noncoding Variants with Deep Learning–Based Sequence Model. *Nature Methods*, *12*(10), 931-934.

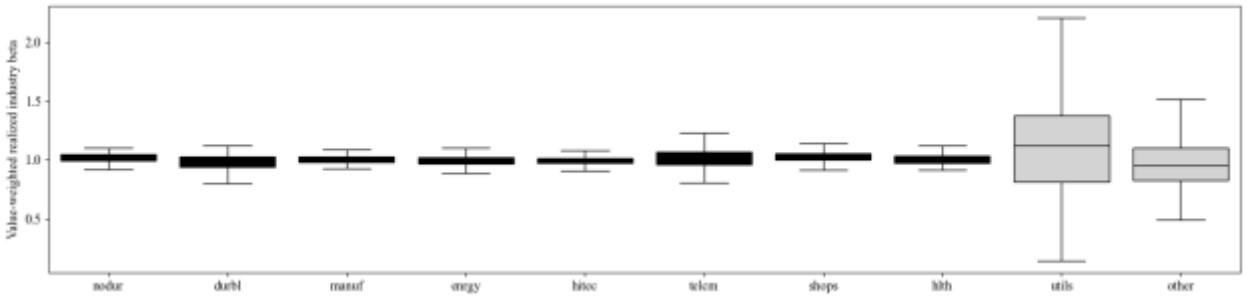
Figure 1
Cross-sectional dispersion of realized industry betas and industry dynamics

This figure presents an overview of the distribution of realized industry betas. Panel A illustrates the cross-sectional dispersion of realized industry betas across all sample months. For each of the ten Fama-French industries, the panel displays two boxplots reflecting the dispersion of beta estimates for industry constituents (black bars) as well as all non-constituents (gray bars) and summarizes the interquartile range (IQR), which spans from the 25th percentile to the 75th percentile. Panel B reports the time-series dispersion of value-weighted realized industry betas. For each month, the industry beta is computed as the value-weighted average of realized betas from industry constituents. Panel C presents the evolution of market-capitalization-weighted industry allocations across all sample months. The sample comprises all firms that were listed on the NYSE, AMEX, or NASDAQ at any point between April 1970 and December 2023.

Panel A: Cross-sectional dispersion of realized industry betas



Panel B: Time-series dispersion of value-weighted realized industry betas



Panel C: Evolution of market-capitalization-weighted industry allocations

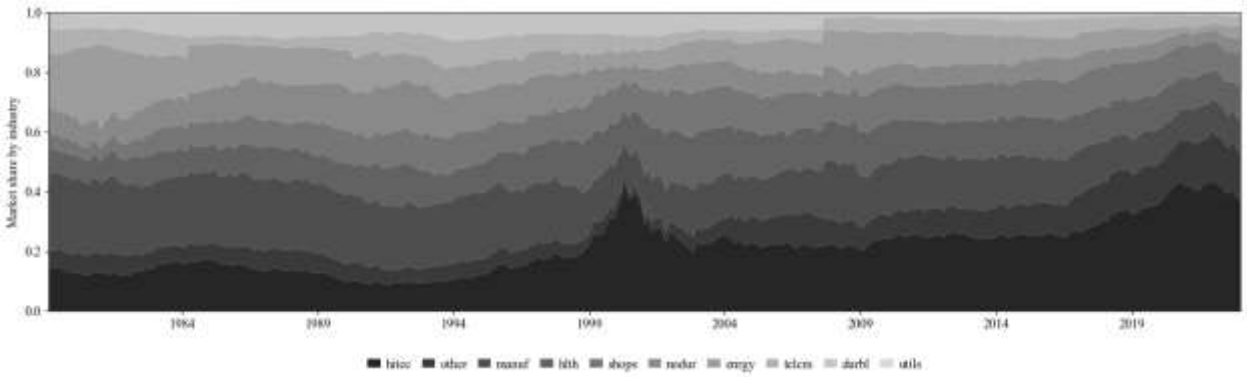
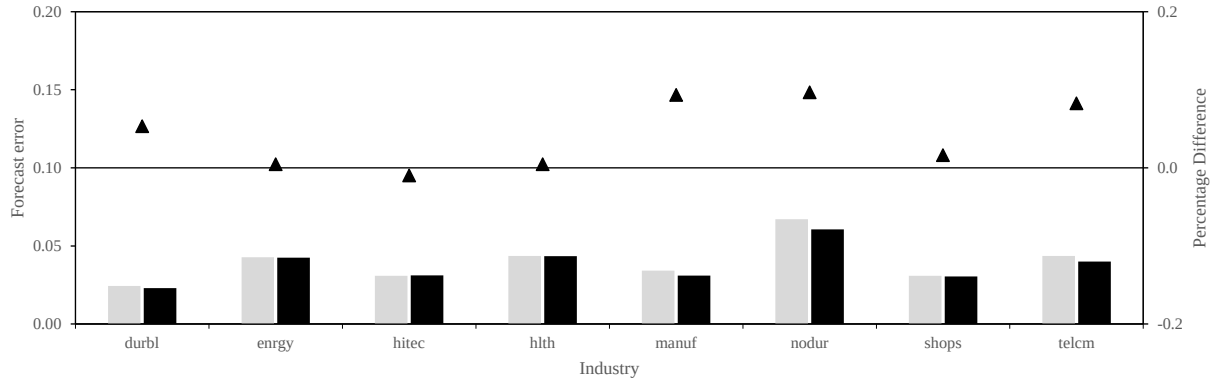


Figure 2
Forecast error by industry and portfolio sorts

This Figure compares the forecast errors for industry beta estimates of the *sep* model (gray bars) and *sim* (black bars) model introduced in Section 4. Beta estimates are computed for eight industries – *durbl, enrgy, hitec, hlth, manuf, nodur, shops, and telcm* – based on the SIC-based industry classification of Fama and French (1997). Forecast errors are defined as time-series averages of monthly value-weighted cross-sectional mean squared errors, computed as $MSE^{(j)}_{t+k|t} = \sum_{i=1}^{N_t} w_{i,t}(\beta^R_{ij,t+k} - \beta^F_{ij,t+k|t})^2$. Realized industry betas are measured over the subsequent twelve months following the forecast month. The figure further reports the percentage differences in average forecast errors (black filled triangles), calculated as one minus the ratio of the average MSE of *sim* to the average MSE of *sep*. Panel A presents results based on the full cross-section, while Panel B shows the time-series averages of forecast errors for portfolios sorted into deciles by predicted beta levels at the end of each month. Stocks are independently sorted for each forecasting model. The sample comprises all firms that were listed on the NYSE, AMEX, or NASDAQ at any point between April 1970 and December 2023.

Panel A: Forecast errors by industry



Panel B: Forecast errors by decile portfolios

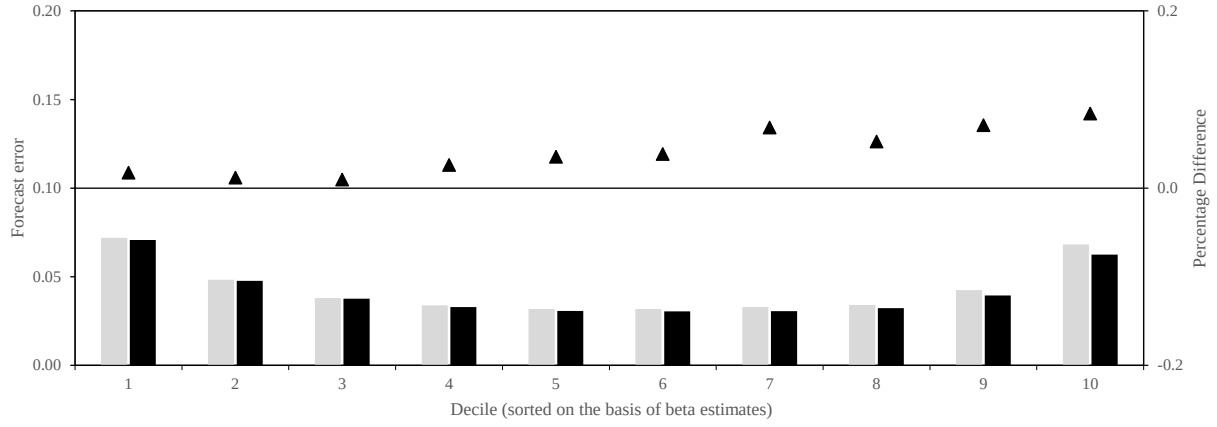
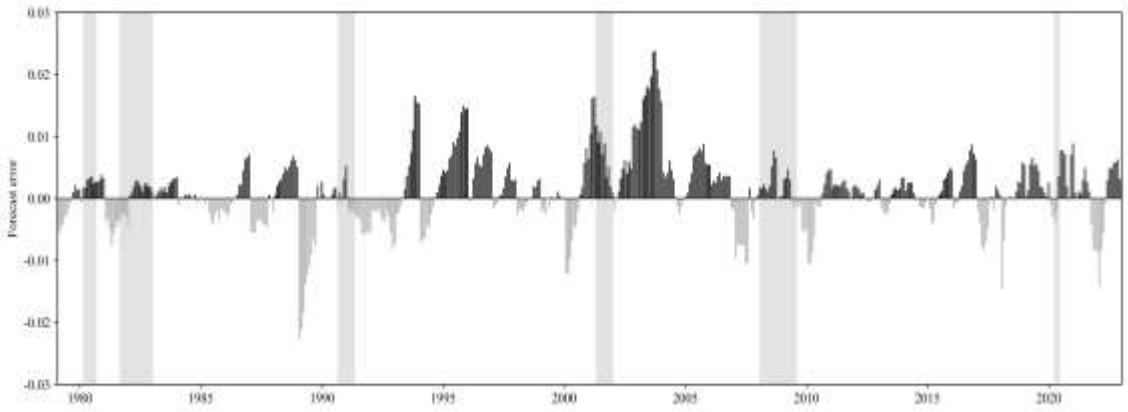


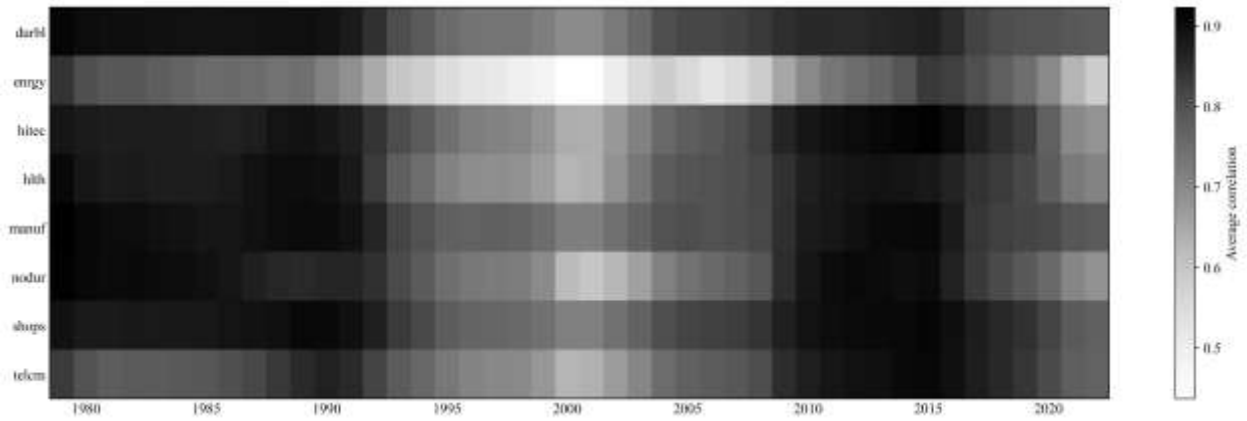
Figure 3
Forecast error over time

This figure illustrates the time-series dynamics of forecast errors for industry beta estimates. Beta estimates are computed for eight industries – *durbl, enrgy, hitec, hlth, manuf, nodur, shops, and telcm* – based on the SIC-based industry classification of Fama and French (1997). Panel A plots the differences in absolute forecast errors. For each month, differences are computed between the absolute value-weighted MSE of *sep* and that of *sim* for each industry and then averaged across all industries. A positive forecast error indicates superior performance of the *sim* model (black bars), while negative forecast errors indicate a lower forecast error for the *sep* model (gray bars). NBER recession periods are shaded in gray. For each industry, Panel B shows the average correlation of the target variables, i.e., the average correlation each industry exhibits with all other industries over time. Panel C reports the fraction of years in the out-of-sample period during which each model is included in the Model Confidence Set (MCS) of Hansen, Lunde, and Nason (2011). The black bars represent the *sim* model, while the gray bars represent the *sep* model. The sample comprises all firms that were listed on the NYSE, AMEX, or NASDAQ at any point between April 1970 and December 2023.

Panel A: Differences in forecast errors over time



Panel B: Average correlation of target variables



Panel C: Inclusion in Model Confidence Set

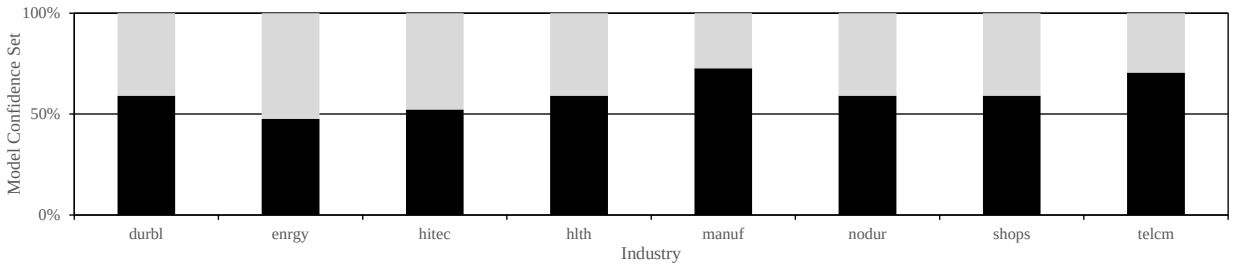


Table 1
Mean squared errors of industry beta estimates

This table reports the mean squared errors (MSE) for industry beta forecasts generated by the machine learning and benchmark models introduced in Section 4.2 and 4.4, respectively, and summarized in Appendix Table A2. MSE are computed as $MSE_{t+k|t}^{(j)} = \sum_{i=1}^{N_t} w_{i,t}(\beta_{ij,t+k}^{R} - \beta_{ij,t+k|t}^{F})^2$ and aggregated across time and industries, either using market-capitalization-based (value-weighted) or equal-weighted schemes (Panel A). Beta estimates are computed for eight industries – *durbl, enrgy, hitec, hlth, manuf, nodur, shops, and telcm* – based on the SIC-based industry classification of Fama and French (1997). Panel B reports the fraction of months in which the column model achieves a lower value-weighted average forecast error relative to the model in the corresponding row. The sample comprises all firms that were listed on the NYSE, AMEX, or NASDAQ at any point between April 1970 and December 2023.

	Benchmark estimators					ML models	
	ols_5y_m	*ols_1y_d*	*ewma_s*	*ewma_l*	*bsw*	*sep*	*sim*
Panel A: Average Forecast Errors							
MSE, value-weighted (%)	14.62	8.10	8.03	7.92	7.30	3.97	3.78
MSE, equal-weighted (%)	37.67	18.04	19.09	18.07	14.58	7.32	6.86
Panel B: Average Forecast Errors over Time							
Benchmark estimators							
vs. ols_5y_m		88.15	89.20	88.85	90.24	99.30	99.65
vs. ols_1y_d	11.85		51.22	64.46	82.58	99.65	99.30
vs. ewma_s	10.80	48.78		58.89	70.73	97.56	96.86
vs. ewma_l	11.15	35.54	41.11		73.52	98.61	97.91
vs. bsw	9.76	17.42	29.27	26.48		97.91	97.56
ML models							
vs. sep	0.70	0.35	2.44	1.39	2.09		63.83
vs. sim	0.35	0.70	3.14	2.09	2.44	36.17	

Table 2
Anomaly portfolio hedging

This table shows the market-neutral anomaly portfolios based on industry beta forecasts that were obtained from the models described in Section 6.1. The anomaly considered are firm size (*me*), book-to-market (*bm*), and illiquidity (*illiq*). Beta estimates are computed for eight industries – *durbl, enrgy, hitec, hlth, manuf, nodur, shops, and telcm* – based on the SIC-based industry classification of Fama and French (1997). Portfolio optimization follows the optimization framework also detailed in Section 6.1. The table entries are the time-series averages of the ex-post industry betas (β) for the long–short portfolios (LS) sorted on each anomaly variable. The associated *t*-statistics, shown in parentheses, are computed using Newey and West (1987) robust standard errors with eleven lags. The null hypothesis is that the respective industry beta for each long-short portfolio is equal to zero. The sample comprises all firms that were listed on the NYSE, AMEX, or NASDAQ at any point between April 1970 and December 2023.

		durbl	enrgy	hitec	hlth	manuf	nodur	shops	telcm
	Model	β_{LS}	β_{LS}	β_{LS}	β_{LS}	β_{LS}	β_{LS}	β_{LS}	β_{LS}
bm	sep	(0.01)	0.01	(0.02)	(0.06)	(0.01)	(0.06)	(0.02)	(0.01)
		(-1.05)	***(1.19)***	*(-4.15)*	*(-13.34)*	*(-2.06)*	*(-7.60)*	*(-4.38)*	***(-1.16)***
	sim	0.01	0.01	(0.01)	(0.04)	0.01	(0.04)	(0.00)	0.01
		(1.42)	*(2.09)*	*(-2.58)*	*(-10.41)*	***(1.66)***	*(-5.67)*	***(-1.01)***	***(1.04)***
me	sep	0.02	0.03	0.01	0.05	0.03	0.07	0.02	0.02
		(3.97)	*(6.29)*	*(2.83)*	*(8.17)*	*(4.76)*	*(6.32)*	*(4.24)*	*(3.44)*
	sim	0.00	0.02	(0.00)	0.03	0.01	0.05	0.00	0.00
		(0.26)	*(3.55)*	***(-0.03)***	*(6.34)*	*(1.82)*	*(5.27)*	***(0.59)***	***(0.39)***
illiq	sep	(0.02)	(0.04)	(0.02)	(0.05)	(0.03)	(0.07)	(0.02)	(0.03)
		(-4.60)	*(-7.01)*	*(-4.39)*	*(-9.50)*	*(-4.43)*	*(-6.68)*	*(-3.99)*	*(-4.99)*
	sim	(0.01)	(0.02)	(0.01)	(0.03)	(0.00)	(0.05)	(0.00)	(0.01)
		(-1.23)	*(-4.14)*	***(-1.56)***	*(-7.36)*	***(-0.59)***	*(-5.42)*	***(-0.04)***	*(-1.72)*

Table 3
Minimum variance portfolios

This table summarizes the characteristics of minimum variance portfolios obtained from the models described in Section 6.2. In the portfolio optimization procedure, we follow Cosemans et al. (2017) and impose a single-factor structure on the covariance matrix of stock returns, making the estimated betas the primary determinants of the portfolio weights. Beta estimates are computed for eight industries – *durbl, enrgy, hitec, hlth, manuf, nodur, shops, and telcm* – based on the SIC-based industry classification of Fama and French (1997). Each month, we obtain portfolio weights that minimize the expected variance, subject to three constraints: (i) individual stock weights must be positive, (ii) no single stock may exceed a 5% weight, and (iii) the portfolio weights must sum to one. Forecasts of market and idiosyncratic variances are based on daily returns over the preceding twelve months. There are three distinct variations of this optimization procedure as outlined in Section 6.2: First, we optimize portfolio weights based solely on market beta estimates (*market beta*). Second, we extend the model by incorporating idiosyncratic industry risk through the inclusion of an industry dummy variable (*market beta + industry dummy*). Third, we capture more nuanced industry exposure by considering industry betas for each stock derived from the multi-output approach (*market beta + industry betas*). The following risk and return metrics are reported: *Std* denotes the ex-post time-series standard deviation of returns, *Dwnd* the downside deviation (computed over months with negative returns), *Min* the lowest monthly excess return, and *MaxDD* the maximum drawdown from a peak to trough considering multi-month periods. The sample comprises all firms that were listed on the NYSE, AMEX, or NASDAQ at any point between March 1970 and December 2023 and have a market capitalization above the 20th percentile of NYSE stocks.

	Minimum Variance Portfolios			
Optimization procedure	Std [%]	Dwnd [%]	Min [%]	MaxDD [%]
Market beta model	12.07	10.07	-27.63	33.28
Market beta + industry dummy model	11.90	9.98	-27.04	31.38
Market beta + industry beta (*sim*) model	11.32	9.38	-26.51	29.00

Table A1

Table A1
Variable descriptions and definitions

This table shows a detailed summary of the 80 predictors used in the empirical analysis following Drobetz et al. (2024). The sample comprises all firms that were listed on the NYSE, AMEX, or NASDAQ at any point between April 1970 and December 2023. Data is aggregated on a monthly frequency and denominated in U.S. dollars when currency-related. To avoid survivorship bias, we assume that firm-level fundamentals are available four months after fiscal year end, while market data becomes available immediately.

#	Predictor	Description	Definition
Predictors based on accounting information			
1	age	Age	Number of years a firm is included in CRSP
2	at	Total assets	Book value of total assets
3	bm	Book-to-market ratio	Ratio of book and market value of equity
4	capturn	Capital turnover	Ratio of net sales to lagged book value of total assets
5	divpay	Dividend payout ratio	Ratio of dividends paid during the last fiscal year to net income of the last fiscal year
6	ep_covar	Covariability in earnings	Coefficient estimate in the regression of a stock's earnings-to-price ratio on the market's earnings-to-price ratio (i.e., the value-weighted average of all stocks' earnings-to-price ratios)
7	ep_var	Variablity in earnings	Standard deviation of monthly earnings-to-price ratios over the last three years
8	finlev	Financial leverage	Ratio of book value of total assets to market value of equity
9	fxdcos	Fixed cost of sales	Ratio of selling, general, and administrative expenses plus research and development expenses plus advertising expenses to net sales
10	illiq	Illiquidity	Ratio of monthly return to monthly dollar trading volume
11	intermom	Intermediate momentum	Excess return from month -12 to month -7
12	invest	Investment	Year-on-year growth of book value of total assets
13	ivol	Idiosyncratic volatility	Standard deviation of daily residuals from the Fama and French (1992) three-factor model within the previous months
14	ltrev	Long-term reversal	Excess return from month -36 to month -13
15	me	Size	Market value of equity
16	mom	Momentum	Excess return from month -12 to month -2
17	noa	Net operating assets	Ratio of operating assets minus operating liabilities to book value of total assets
18	opaccr	Operating accruals	Ratio of changes in noncash working capital minus depreciation to book value of total assets
19	oplev	Operating leverage	Ratio of operating costs (i.e., the sum of costs of goods sold and selling, general and administrative expense) to market value of total assets
20	ppe	PPE change-to-assets ratio	Ratio of changes in property, plants, and equipment to lagged book value of total assets
21	prof	Profitability	Ratio of gross profits to book value of equity
Technical indicators			
22	relprc	Relative price	Ratio of previous month's price to its highest daily price during the last year
23	roa	Return on assets	Ratio of income before extraordinary items to book value of total assets
24	roe	Return on equity	Ratio of income before extraordinary items to book value of equity
25	ron	Return on net operating assets	Ratio of operating income after depreciation to lagged net operating assets
26	salestoassets	Sales-to-assets ratio	Ratio of net sales to book value of total assets
27	salestoprice	Sales-to-price ratio	Ratio of net sales to market value of equity
28	SGAtosales	SGA-to-sales ratio	Ratio of selling, general, and administrative expenses to net sales
29	strev	Short-term reversal	Excess return from the previous month
30	to	Turnover	Monthly dollar trading volume
Macroeconomic indicator			
31	dfy	Default spread	Yield differential between Moody's Baa- and Aaa-rated corporate bonds
Predictors based on sample estimates of beta			
42 – 51	ols_3m_d_ind	Short-term beta	Sample estimate of changes in industry beta obtained from rolling regressions using a three-month window of daily returns
52 – 61	ols_1y_d_ind	Medium-term beta	Sample estimate of changes in industry beta obtained from rolling regressions using a one-year window of daily returns

62 – 71	ols_5y_d_ind	Long-term beta	Sample estimate of changes in industry beta obtained from rolling regressions using a five-year window of daily returns
Industry classifiers			
72 - 80	ind	Industry classification	Fama and French (1997) industry classification, resulting in 10-1 = 9 industry dummies

Figure A1

Figure A1
Neural Network Architecture

This figure illustrates the architecture of the *sep* model (left) and the *sim* model (right) introduced in Section 4. Gray circles denote the input layer, with the number of nodes corresponding to the dimension of the predictor vector. Black filled circles denote the output layer: in *sep*, the network forecasts a single industry's beta at a time; in *sim*, the network jointly forecasts betas for all industries. Hidden layers apply the hyperbolic tangent (tanh) activation function element-wise to linear combinations of the previous layer's outputs. Each network type uses either one [32], two [32, 16], or three [32, 16, 8] hidden layers, following the geometric pyramid rule (Masters, 1993) with decreasing numbers of neurons in deeper layers. Arrows represent trainable weight parameters between layers.

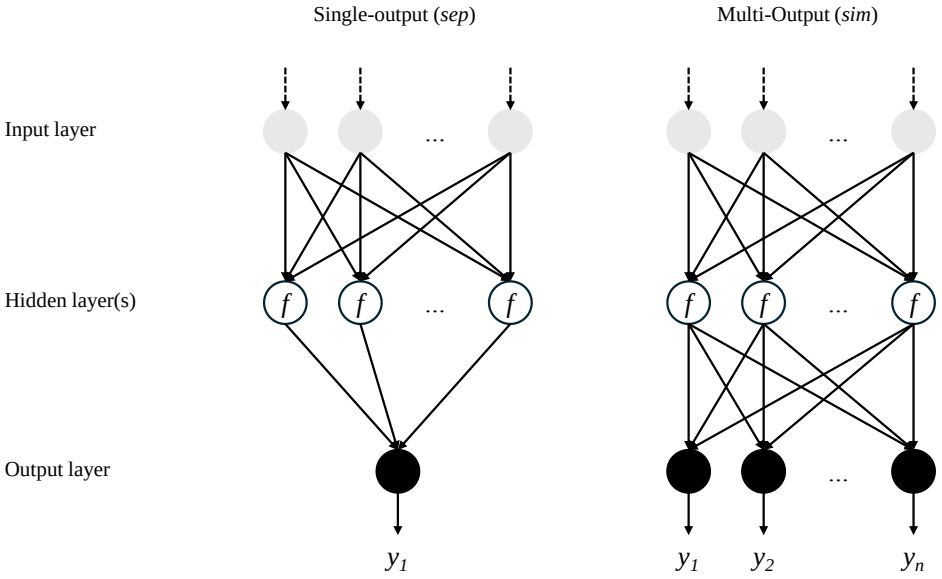


Table A2

Table A2
Forecast models (incl. hyperparameter specifications)

This table documents the hyperparameter settings, network architectures, and additional implementation details for the forecast models introduced in Section 4.

Panel A: Benchmark estimators

Model	Description	Definition
ols_5y_m	Historical beta	Rolling regressions using a five-year window of monthly returns
ols_1y_d	Historical beta	Rolling regressions using a one-year window of daily returns
ewma_s	Exponentially-weighted beta	Rolling regressions using a one-year window of daily returns with exponentially decaying weights (short half-life)
ewma_l	Exponentially-weighted beta	Rolling regressions using a one-year window of daily returns with exponentially decaying weights (long half-life)
bsw	Slope-winsorized beta	Rolling regressions using a one-year window of *winsorized* daily returns

Panel B: Machine learning estimators

#	Hyperparameter	Specification	Definition
General model set up			
	activation	*tanh*	Activation function
	$size_{batch}$	50	Batch size
	$number_{epochs}$	100	Number of epochs
	patience	5	Number of iterations during which the value-weighted MSE is allowed to increase in the validation sample
	dropout rate	0.05	Fractional rate of input variables that are randomly set to zero at each iteration
	batch_normalization	✓	Batch normalization is applied after the last hidden layer to stabilize and accelerate training by standardizing layer inputs.
	learning_rate	0.0001	Learning rate used by the Adam optimizer to update network weights during backpropagation
	l_1, l_2	0.00001	Regularization parameters that penalize large model weights to prevent overfitting.
	ensemble	5	Number of independent seeds used for each specification family at each re-estimation date
Network architecture			
	nn_1	[32]	Architecture uses either one [32], two [32, 16], or three [32, 16, 8] hidden layers, following the geometric pyramid rule (Masters, 1993) with decreasing numbers of neurons in deeper layers
	nn_2	[32, 16]	
	nn_3	[32, 16, 8]	
Single-Output Neural Net (sep)			
	y_i	1	Output node(s)
Multi-Output Neural Net (sim)			
	y_i	8	Output node(s)