cemfi

working paper

2411

# Portfolio management with big data

Francisco Peñaranda
Enrique Sentana

June 2024

CENTRO DE ESTUDIOS
MONETARIOS Y FINANCIEROS
Casado del Alisal 5, 28014 Madrid, Spain
www.cemfi.es

CEMFI Working Paper No. 2411
June 2024

# Portfolio management with big data

**Abstract**

The purpose of this survey is to summarize the academic literature that studies some of the ways in which portfolio management has been affected in recent years by the availability of big datasets: many assets, many characteristics for each of them, many macro predictors, and various sources of unstructured data. Thus, we deliberately focus on applications rather than methods. We also include brief reviews of the financial theories underlying asset management, which provide the relevant background to assess the plethora of recent contributions to such an active research field.

JEL Codes: G11, G12, C55, G17.

Keywords: Conditioning information, intertemporal portfolio decisions, machine learning, mean-variance analysis, stochastic discount factors.

Francisco Peñaranda
Queens College CUNY
francisco.penaranda@qc.cuny.edu

Enrique Sentana
CEMFI
sentana@cemfi.es

**Acknowledgement**

Prepared for Peña, D. (ed.) Nuevos métodos de investigación económica con datos masivos, Fundación Ramón Areces, Madrid, 2025. Sentana gratefully acknowledges financial support from the Santander Research chair at CEMFI.

# 1 Introduction

Big data, machine learning and artificial intelligence are increasingly popular concepts frequently treated as close substitutes in the media. And although it is true that artificial intelligence often relies on machine learning methods, which in turn are trained on big datasets, they reflect different concepts. The focus of our survey is big data. Specifically, we attempt to summarize the academic literature that studies some of the ways in which portfolio management has been affected in recent years by the availability of large datasets. These include prices and returns for many assets, as well as many characteristics for each of them, multiple macro predictors and their non-linear transformations, and various sources of unstructured data, including texts and photos, and potentially audios and videos. Therefore, we focus on applications rather than methods, which we do not review in detail.

Readers interested in a deeper understanding of some of the machine learning methods employed in the different applications that we discuss could read the references in the papers that we review, as well as the recent surveys and books we mention next, which cover different aspects. For example, Israel, Kelly and Moskowitz (2020) emphasize that finance is different from the typical machine learning application because of smaller datasets, low signal-to-noise ratios, non-stationary environments, and the need for interpretability. In turn, Bartram, Branke and Motahari (2020), Mirete-Ferrer et al. (2022) and Guidolin (2024) focus on asset management, Giglio, Kelly and Xiu (2022) on asset pricing models, while Kelly and Xiu (2023) review empirical finance applications more widely, including the construction of optimal portfolios.

In terms of books, López de Prado (2018) and Dixon, Halperin and Bilokon (2020) can serve as an introduction to machine learning for readers with a background in financial econometrics, while the main focus of Guida (2019), Capponi and Lehalle (2023) and Cao (2023) is investment applications of big data and machine learning. Given that dependence is an important property of financial data, Peña and Tsay (2021) is also useful for its pedagogical review of statistical learning methods applied to big dependent datasets.

Like many of those references, we must emphasize that the concept of big data in the context of asset management applications is relative to the size of the typical datasets available at the end of the twentieth century. Consequently, with the exception of the alternative data sources that we discuss in the last part of our survey, the datasets that we refer to are often noticeably smaller and more structured than those regularly analyzed in other data science applications.

In line with most of the literature, our review covers mainly stock market applications, although it also discusses some other asset classes, such as fixed-income, foreign exchange and commodities. Moreover, we focus on either cash assets or forwards and futures rather than

options and other financial derivatives (see chapter 6 in Hull (2021) and section 3 of Bartram, Branke and Motahari (2020) for additional references). Given our interest in asset management, we also exclude many papers whose only objective is the mere forecasting of asset returns or prices, as well as those whose scope is asset pricing but ignoring its potential portfolio management implications.

In the interest of space, we by and large ignore two important areas that have been revolutionized in the twenty first century by the availability of high frequency data: realized volatility and correlation (see Barndorff-Nielsen and Shephard (2007), Andersen, Bollerslev and Diebold (2010) and Aït-Sahalia and Jacod (2014)), and algorithmic trading through execution management systems that efficiently process trading orders by minimizing both standard transaction costs and market impact (see López de Prado (2020) and chapters 7 and 8 of Cao (2023)). We have also excluded a review of the impact on portfolio evaluation of the massive growth in the number of mutual funds and other investment vehicles that have become available in the last few decades, whose main consequence is the so-called multiple testing problem (see Barras, Scaillet and Wermers (2010, 2018), Harvey and Liu (2020) and Giglio, Liao and Xiu (2021)).

To provide the relevant background to the plethora of recent contributions to this incredibly active field of research, we include brief reviews of the financial theories underlying asset management in practice. In particular, we study the relationship between mean-variance analysis and stochastic discount factors to introduce the reader to the terminology that pervades the modern empirical finance literature. We do so both when investors rely exclusively on the unconditional distribution of asset returns in making their decisions and when they exploit other sources of information at their disposal at the time they design their trading strategies. Finally, we also present the theory of intertemporal portfolio decisions, which is relevant when important aspects of the investment opportunity set are changing over time in predictable ways.

The rest of the survey is organized as follows. First, in section 2 we consider a large cross-section of assets in an unconditional set up. Then, we extend our analysis to conditioning information in section 3, considering situations in which investors exploit a large number of asset characteristics or macroeconomic indicators in constructing their portfolios. Finally, section 4 discusses two areas that are still relatively small, but which offer substantial research potential: intertemporal portfolio decisions using big data, and what one may call alternative data.

# 2 Static mean-variance frontiers and stochastic discount factors

## 2.1 Theoretical background

Mean-variance (MV) analysis is widely regarded as the cornerstone of modern investment theory. Despite its simplicity, and the fact that over seven decades have elapsed since Markowitz's (1952) seminal work on the theory of portfolio allocation under uncertainty, it remains the most widely used asset allocation method. There are several reasons for its popularity. First, it provides a very intuitive assessment of the relative merits of alternative portfolios, as their risk and expected return characteristics can be compared in a two-dimensional graph. Second, the portfolios on the MV frontiers defined below are spanned by two funds at most, a property that simplifies their calculation and interpretation, and that also led to the derivation of the Capital Asset Pricing Model (CAPM). Third, MV analysis becomes the natural approach if we assume Gaussian or elliptical distributions for asset returns because in that case it is fully compatible with expected utility maximization regardless of investor preferences. Finally, MV frontiers for returns are intimately related to MV frontiers for stochastic discount factors (SDFs) put forward by Hansen and Jagannathan (1991), which represented a major breakthrough in the way financial economists look at data on asset returns to discern which asset pricing theories are not empirically falsified.

For simplicity, we focus on the case in which all investable assets have zero costs and there are no restrictions on long or short positions. Let us denote by $\mathbf{r} = (r_1, \ldots, r_i, \ldots, r_n)'$ the vector of $n$ excess returns available to investors. Although zero cost investments with arbitrary short and long positions might seem totally unrealistic in practice, $r_i$ usually corresponds to the payoffs to a simple trading strategy implemented by buying or selling futures contracts in organized markets. As a result, investors can freely scale up or down the payoffs to each of those strategies. A related advantage is that there is no need to impose a cost constraint because the initial payoff of taking those positions is 0, and often any margin requirements imposed by the market in which the futures contracts are traded will be rewarded at the safe interest rate.

Formally, the MV principle consists in choosing the weights of a portfolio that minimize the standard deviation of its excess returns for each possible target level of expected returns. If the first two moments of returns were known, then it would be straightforward to apply Markowitz's optimal portfolio formulas. In practice, of course, the mean vector and covariance matrix of $\mathbf{r}$ are unknown, and the sample mean, standard deviations and correlations of its elements only provide noisy estimators of the required theoretical quantities.

There are several ways of computing the optimal MV weights. We find it convenient to adopt

the Generalized Method of Moments (GMM) framework in Peñaranda and Sentana (2011), which is effectively equivalent to Gaussian pseudo maximum likelihood (PML) estimation in this context. The most natural possibility would be to implicitly define the vector of risk premia $\boldsymbol{\mu}$ as

$$E(\mathbf{r}-\boldsymbol{\mu})=\mathbf{0}$$

and either the second moment matrix $\boldsymbol{\Gamma}$ as

$$E(\mathbf{r}\mathbf{r}'-\boldsymbol{\Gamma})=\mathbf{0},$$

or the covariance matrix $\boldsymbol{\Sigma}$ as

$$E[\mathbf{r}(\mathbf{r}-\boldsymbol{\mu})'-\boldsymbol{\Sigma}]=\mathbf{0}.$$

Naturally, for MV analysis to make sense, we need to assume that $\mathbf{r}$ belongs to the collection of random variables defined on the underlying probability space with bounded second moments, which is equivalent to assuming that (the norm of) $\boldsymbol{\Gamma}$ is bounded. As a result, $\boldsymbol{\mu}$ will be well-defined and $\boldsymbol{\Sigma}$ will be bounded. We also assume for simplicity that the smallest eigenvalue of $\boldsymbol{\Sigma}$ is strictly positive, which implies that none of the zero-cost portfolios in $\mathbf{r}$ is either riskless or redundant, and moreover, that it is not possible to generate a riskless portfolio from $\mathbf{r}$ other than the trivial one. With these assumptions, $\boldsymbol{\Gamma}$ is invertible, and we can compute the optimal MV portfolio weights as

$$\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}=(1+\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},$$

where we have exploited the Sherman-Morrison-Woodbury formula, which says that

$$\boldsymbol{\Gamma}^{-1}=(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}')^{-1}=\boldsymbol{\Sigma}^{-1}-(1+\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}. \tag{1}$$

Strictly speaking, though, estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$ or $\boldsymbol{\Sigma}$ is unnecessary, as it can be avoided by resorting to the concept of mean representing portfolios introduced by Chamberlain and Rothschild (1983), which in their uncentred and centred versions are such that $E(\mathbf{r})=E(\mathbf{r}a^{\circ})=cov(\mathbf{r},\eth^{\circ})$, respectively. Consequently,

$$\begin{aligned} a^{\circ} &= E(\mathbf{r}')[E(\mathbf{r}\mathbf{r}')]^{-1}\mathbf{r}, \\ \eth^{\circ} &= E(\mathbf{r}')[V(\mathbf{r})]^{-1}\mathbf{r}=[1-E(a^{\circ})]^{-1}a^{\circ}. \end{aligned}$$

On this basis, one can construct the arbitrage (i.e. zero-cost) MV frontier by scaling the mean representing portfolios, so that its elements will be of the form

$$r^{MV}(\mu)=\mu\frac{1}{E(a^{\circ})}a^{\circ}=\mu\frac{1}{E(\eth^{\circ})}\eth^{\circ},$$

being thus defined for all possible values of the target expected return $\mu$ as long as $E(\mathbf{r})$ is not proportional to a vector of $n$ ones, $\boldsymbol{\iota}_n$, as it would happen in equilibrium if there was a single risk neutral investor. As a result,

$$V[r^{MV}(\mu)] = \frac{1 - E(a^\circ)}{E(a^\circ)}\mu^2 = \frac{1}{E(\eth^\circ)}\mu^2.$$

Consequently, the function $\{\mu, V[r^{MV}(\mu)]\}$ will be a parabola tangent to the origin in MV space, while the related function $\{\mu, \sqrt{V[r^{MV}(\mu)]}\}$ will be a reflected straight line in mean-standard deviation space. In addition, the weights of the different assets in $r^{MV}(\mu)$ are proportional to their weights on the mean representing portfolios in view of the previous expressions.

Aside from those weights, the only other unknown parameter that matters is

$$\theta = E(\eth^\circ) = [1 - E\,(a^\circ)]^{-1}\,E\,(a^\circ) = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},$$

which we can interpret as the maximum (square) Sharpe ratio attainable. The Sharpe ratio, defined as the ratio of the expected excess return of an investment to its standard deviation, is one of the most common measures used by financial market practitioners to rank fund managers and to evaluate the attractiveness of investment strategies in general (see Sharpe (1994)). Apart from its simplicity, and the fact that it is a rather natural risk-adjusted measure of performance, it has also the convenient property of being numerically invariant to the degree of leverage of the position. However, the Sharpe ratio is not without limitations, as the academic literature on performance evaluation has made clear (see for example Goetzmann et al. (2007)).

In turn, modern asset pricing theories are written in terms of SDFs, which are univariate random variables typically denoted by $m$ that transform asset payoffs into asset costs by discounting them differently in different states of the world. In this context, the elements of the SDF MV frontier based on arbitrage portfolios only are the univariate random variables of minimum variance among those that correctly price both $\mathbf{r}$ and a fictitious safe asset with return $1/c$ as $E[m(\mathbf{r}',1/c)] = (\mathbf{0}', 1)$. Hansen and Jagannathan (1991) show that those elements will be given by

$$m^{MV}(c) = \frac{c}{1 - E(a^\circ)}(1 - a^\circ) = c\{1 - [\eth^\circ - E(\eth^\circ)]\},$$

so that they are also spanned by a single "fund". In turn, their variance will be

$$Var[m^{MV}(c)] = c^2\frac{E(a^\circ)}{1 - E(a^\circ)} = c^2 E(\eth^\circ),$$

which is a perfect square in $c$ that depends on the reciprocal of the same single parameter $\theta$. This confirms the duality between the SDF and portfolio frontiers because the maximum (squared) Sharpe ratio is equal to $Var[m^{MV}(c)]/c^2$. Once again, the function $\{c, Var[m^{MV}(c)]\}$

is a parabola in MV space, while $\{c, \sqrt{Var[m^{MV}(c)]}\}$ is a half line starting from the origin in mean-standard deviation space.

Given a vector of $n$ excess returns $\mathbf{r}$, we can estimate both frontiers from the following exactly identified system of $n+1$ moment conditions:

$$E\left(\begin{array}{c}\mathbf{r}\mathbf{r}'\boldsymbol{\phi}^{\circ}-\mathbf{r}\\ \mathbf{r}'\boldsymbol{\phi}^{\circ}-\mu^{\circ}\end{array}\right)=\mathbf{0}, \tag{2}$$

where $\mu^{\circ}=\theta/(1+\theta)=1/(1+\eta)$, with $\eta=\theta^{-1}$, identifies $E\left(a^{\circ}\right)=E\left(a^{\circ 2}\right)$, and $\boldsymbol{\phi}^{\circ}$ the portfolio weights of this uncentred mean representing portfolio. Under standard regularity conditions, the resulting GMM estimator of $\theta$ will converge in probability to its true value, and the same applies to the weights of the mean representing portfolios in $\boldsymbol{\phi}^{\circ}$. Therefore, the GMM estimators of $V[r^{MV}(\mu)]$ will also converge in probability to their population counterparts for fixed $\mu$. Further, we can easily show that the GMM estimators of the entire arbitrage and SDF MV frontiers will converge uniformly to their population analogues over any finite range.

Despite the uniform consistency, though, the arbitrage and SDF MV frontiers are subject to substantial sample variability. This variability is so important that if one does not take it into account, one would form very different optimal portfolios depending on the sample, and more importantly, would reach rather different conclusions about the available risk-return trade-offs.

In particular, there is a clear tendency to reach overly optimistic conclusions about the MV trade-offs that investors really face in the future (out-of-sample) by relying on statistics obtained from historical data (in-sample), to the extent that regulators force investment managers to systematically add the hackneyed caveat "past performance is not indicative of future results" to their marketing material. In this respect, Kan and Zhou (2007) proved that when $\mathbf{r}$ is identically and independently distributed (*i.i.d.*) as a multivariate normal, the finite sample distribution of the maximum square Sharpe ratio estimator follows a non-central distribution whose probability mass is asymmetrically distributed above the true value.[1] For that reason, they suggest replacing the unrestricted sample estimator of $\theta$, $\hat{\theta}_T$, with the following unbiased estimator

$$\dddot{\theta}=\frac{(T-n-2)\hat{\theta}_T-n}{T}. \tag{3}$$

For a fixed, relatively small number of assets, $n$, the main source of the problem is the estimation error in expected returns, whose effect is first-order (see Best and Grauer (1991) for a study of the sensitivity of the MV weights to changes in the means). As Merton (1980) showed when (log) asset prices follow diffusions with constant drift and instantaneous covariance matrix,

[1] In fact, Sentana (2009) shows that in the case in which the true Sharpe ratio is 0, the distribution of the estimated one will only have probability mass on the positive side.

standard deviations and correlations will be estimated more and more precisely as the number of observations increases for a fixed time span when one increases the observation frequency, while the precision of expected returns only increases if the sample span gets larger.

## 2.2 Big data on cross-sections of assets

The estimation methods discussed in the previous section are designed for situations in which the number of assets $n$ is substantially smaller than the number of observations $T$. When the number of assets is large, though, the sampling uncertainty of the portfolio weights increases even further. More importantly, the maximum Sharpe ratio in the sample can never decrease when the number of assets grows. In fact, if there are at least as many assets as time series observations, the sample covariance matrix will become singular, the portfolio weights underidentified, and the law of one price seemingly violated as the maximum Sharpe ratio will become infinity.

At first sight, this might seem a largely irrelevant question. However, as its name indicates, the S&P 500, arguably the world's most famous stock market index, combines the prices of five hundred assets traded in the New York Stock Exchange (NYSE). Similarly, the Russell 3000, another capitalization-weighted stock market index that represents approximately 97% of the equity value of all publicly traded companies in the U.S. stock market, measures the performance of the three thousand largest publicly held companies incorporated in the country.[2] Three thousand observations correspond to roughly eleven and a half years of daily data, almost 57 years and nine months of weekly data, and 250 years of monthly data. That means we would necessarily encounter a singular covariance matrix when working with monthly data even though trading in the NYSE started 232 years ago in 1792. For academic research purposes, the main source of data is the University of Chicago Center for Security Prices (CRSP) US stock database, which contains daily and monthly data for over 32,000 active and inactive securities with primary listings on the most important US stock markets, including among others NYSE, NYSE American and NASDAQ.

Next, we discuss three alternative approaches that have been suggested to reduce the effects of sampling variability.

### 2.2.1 Parametric restrictions

Although expression (2) shows that the estimation of the portfolio weights does not require the estimation of the mean vector or the covariance matrix of the vector of asset returns, it is

[2] As a comparison, the MSCI All Countries World Index (ACWI), with 2,840 constituents, represents 85% of the global investable equity opportunity set.

fairly straightforward to impose empirically plausible restrictions on the elements of these two moments to reduce the sampling variability of the optimal weights. Given that the main source of uncertainty arises because of the estimation of expected excess returns, whose precision depends on the length of the sample span used rather than the frequency of observations, it seems natural to impose restrictions on risk premia, such as those derived from an asset pricing model.

In principle, nothing prevents us from considering non-linear models for the SDF, $m$. Nevertheless, the standard approach in empirical finance is to model it as an affine transformation of some $k \leq n$ observable risk factors $\mathbf{f}$, even though this ignores that $m$ must be positive with probability 1 to avoid arbitrage opportunities. Popular risk factors are either returns on traded portfolios, like the market portfolio in the CAPM, or macro-variables, such as the per capita labour income used as a proxy variable for the returns to human capital by Jagannathan and Wang (1996).

In this context, we can express the pricing equation as

$$E[(\lambda_0 - \boldsymbol{\lambda}'\mathbf{f})\mathbf{r}] = \mathbf{0} \tag{4}$$

for some real numbers $(\lambda_0, \boldsymbol{\lambda}')'$. Although $\mathbf{r}$ only contains assets with 0 cost, which leaves the *scale* and *sign* of $m$ undetermined, we would like our candidate SDF to price other assets with positive prices. Therefore, we require a *scale normalization* to rule out the trivial solution $(\lambda_0, \boldsymbol{\lambda}')' = (0, \mathbf{0}')'$. Assuming for simplicity that $\mathbf{f}$ and $\mathbf{r}$ do not share any common elements, we can add the pricing conditions (4) to the exactly identified moment conditions (2) that define the uncentred mean representing portfolio, thereby obtaining

$$E\begin{bmatrix} \mathbf{r}(1 - \mathbf{f}'\boldsymbol{\lambda}) \\ \mathbf{r}\mathbf{r}'\boldsymbol{\phi}^\circ - \mathbf{r} \\ \mathbf{r}'\boldsymbol{\phi}^\circ - \mu^\circ \end{bmatrix} = \mathbf{0},$$

where the unknown parameters are $(\boldsymbol{\lambda}', \boldsymbol{\phi}^{\circ\prime}, \mu^\circ)$. Thus, we can obtain more efficient estimators of the MV weights that exploit the pricing equations.

Let us partition $\mathbf{r}$ into two sets of portfolios $\mathbf{r}_1$ and $\mathbf{r}_2$ of dimensions $n_1$ and $n_2$, respectively, with $n = n_1 + n_2$, so that $\mathbf{r}' = (\mathbf{r}_1', \mathbf{r}_2')$. A closely related alternative is to impose MV spanning restrictions, which can be regarded as if we assumed a linear factor pricing model in which the pricing factors $\mathbf{f}$ coincide with some excess returns $\mathbf{r}_1$, as in the Fama and French (1993, 2015) models that we discuss in section 3.2. In that case, we say that $\mathbf{r}_1$ spans the zero-cost and SDF MV frontiers generated from $\mathbf{r}_1$ and $\mathbf{r}_2$. Under the null hypothesis that this is the indeed case, there will be only one pair of MV frontiers. Under the alternative, there will be two: the frontiers generated from $\mathbf{r}_1$ alone, and the ones generated from $\mathbf{r}$, which will only touch at the origin.

In the GMM context described in the previous section, the imposition of the null hypothesis of spanning on the weights of the uncentred moment conditions gives rise to the overidentified system

$$E\left[\begin{pmatrix}\mathbf{r}_1\\ \mathbf{r}_2\\ 1\end{pmatrix}\mathbf{r}_1'\boldsymbol{\phi}_1^\circ-\begin{pmatrix}\mathbf{r}_1\\ \mathbf{r}_2\\ \mu^\circ\end{pmatrix}\right]=\mathbf{0}. \tag{5}$$

The optimal GMM estimator of $\theta=\mu^\circ/(1-\mu^\circ)$ obtained from this conditions will generally be more efficient than the corresponding estimator obtained from the unrestricted system (2) as long as the equality restriction $\boldsymbol{\phi}_2^\circ=\mathbf{0}$ holds. Moreover, this estimator will also be generally more efficient than the one obtained from the just identified $n_1+1$ moment conditions

$$E\left[\begin{pmatrix}\mathbf{r}_1\\ 1\end{pmatrix}\mathbf{r}_1'\boldsymbol{\phi}_1^\circ-\begin{pmatrix}\mathbf{r}_1\\ \mu^\circ\end{pmatrix}\right]=\mathbf{0}.$$

Unfortunately, the reduction in sampling uncertainty resulting from this type of restriction is noticeable but not particularly big. In fact, it completely disappears in the case of (5) when the joint distribution of the vector of excess returns $\mathbf{r}$ is elliptical (see Peñaranda and Sentana (2015)). Besides, one runs the risk of misspecification, which would typically imply inconsistencies in the parameter estimators and portfolio weights.

When the number of assets $n$ is large, another popular empirical strategy is to impose some structure on $\boldsymbol{\Sigma}$. A suggestion with a long tradition is an exact $k$ factor structure in which

$$\boldsymbol{\Sigma}=\mathbf{B}\mathbf{B}'+\boldsymbol{\Upsilon}, \tag{6}$$

where $\mathbf{B}$ is $n\times k$ and $\boldsymbol{\Upsilon}$ diagonal, so that the number of parameters to estimate increases with $n$ rather than $n^2$. This covariance model is equivalent to

$$\begin{aligned}\mathbf{r} &= \boldsymbol{\mu}+\mathbf{B}\mathbf{f}+\mathbf{u},\\ \begin{pmatrix}\mathbf{f}\\ \mathbf{u}\end{pmatrix} &\sim D\left[\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix},\begin{pmatrix}\mathbf{I}_k & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Upsilon}\end{pmatrix}\right],\end{aligned}$$

where $\mathbf{I}_k$ is the identity matrix of order $k$ and $D(\mathbf{m},\mathbf{V})$ denotes a distribution with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{V}$. In this context, $\mathbf{B}$ contains the sensitivities of the different assets to the $k$ orthogonal sources of systematic risk in the economy that affect most of them, $\mathbf{f}$, while the diagonal elements of $\boldsymbol{\Upsilon}$ represent the variances of the idiosyncratic risks $\mathbf{u}$, which are uncorrelated across assets. An important special case is the so-called equicorrelated structure in which the correlations between any two assets are assumed identical (see Elton and Gruber (1973)). This assumption is so popular that the Chicago Board of Options Exchange (CBOE)

exploits this restricted single factor model to generate an analogue to the VIX index for implied correlations among the constituents of the S&P 500. Another example is the so-called diagonal or market model in Sharpe (1963), in which the single common factor is the return to the market portfolio and the factor loadings in $\mathbf{B}$ are the market betas multiplied by the standard deviation of the market portfolio. Fan, Fan and Lv (2008) formally compare the precision of the sample covariance matrix with the one obtained with a multifactor version of Sharpe's (1963) market model in large $n$ and $T$ samples. They find that although their observable factor structure does not improve much the estimation of the covariance matrix itself, there are substantial gains in the estimation of its inverse, which is the one that matters for MV portfolio allocation.

In fact, the covariance specification (6) has implications for risk premia when $n$ is large. Specifically, the exact version of Ross (1976) arbitrage pricing theory (APT) implies that

$$\boldsymbol{\mu} = \mathbf{B}\boldsymbol{\pi} \tag{7}$$

for a $k$-dimensional vector $\boldsymbol{\pi}$ when the factors are pervasive so that the smallest eigenvalue of $\mathbf{B}'\mathbf{B}$ diverges as $n$ grows. Therefore, the moment condition (4) holds, but the pricing factors $\mathbf{f}$ must reflect the $k$ common sources of risk in the economy.[3]

Unfortunately, the diagonality of $\boldsymbol{\Upsilon}$ is excessively restrictive unless $k$ is rather large. A possible solution is to use bifactor models, which combine a few pervasive factors with many "sectoral" factors that affect most firms belonging to the same industrial category and are orthogonal across categories (see Aït-Sahalia and Xiu (2017) for an example with high-frequency data).

Nevertheless, Chamberlain and Rothschild (1983) proved that the APT would continue to hold in what they called approximate factor structures, in which $\boldsymbol{\Upsilon}$ is not necessarily diagonal, but its largest eigenvalue remains bounded as $n$ grows, so that a law of large numbers still applies cross-sectionally to the idiosyncratic terms. They also showed that in those circumstances, principal components analysis (PCA), an early example of unsupervised learning that performs a data compression numerically equivalent to factor analysis when $\boldsymbol{\Upsilon}$ is scalar, yields a consistent basis of the space of portfolios that mimic the latent factors $\mathbf{f}$ with vanishing tracking error as the number of asset grow. Not surprisingly, data-based selection procedures for empirically determining the appropriate number of factors when $n$ is large rely on suitably grouping the eigenvalues of $\boldsymbol{\Sigma}$ into $k$ big ones and $n-k$ small ones.

The first papers that used a large cross-section of assets to test the APT were Lehmann and

[3] As Chamberlain (1983) highlighted, strictly speaking the APT only implies that the (Euclidean) norm of the difference between the left- and right-hand sides of (7) remains bounded as the number of asset increases, which complicates its testing (see Dello Preite et al (2024) for recent empirical evidence on this point and Da, Nagel and Xiu (2023) for a related discussion).

Modest (1988) and Connor and Korajczyk (1988). They replaced the latent factors by mimicking portfolios obtained by maximum likelihood and PCA, respectively. In turn, Connor and Korajczyk (1993) constitutes the first attempt to determine the number of factors an approximate factor model requires to fit a large cross-section of stock returns. Subsequently, Bai and Ng (2002) proposed consistent estimators of the number of factors based on a penalized least squares objective function associated to PCA when both the number of assets $n$ and the number of time series observations $T$ go to infinite at the same rate. The main problem, though, is the existence of weak factors, whose associated eigenvalues grow with $n$ but a rather slow pace (see Onatski (2012) for their implications for PCAs).

These procedures, however, ignore the APT restriction (7) by focusing exclusively on (6). In this respect, Lettau and Pelger (2020a) propose a modification of the criterion function in Bai (2003) that penalizes discrepancies from both (6) and (7). Thus, they can detect weak pricing factors with high Sharpe ratios. In Lettau and Pelger (2020b), they apply their methods to the popular Fama and French (1993) cross-section of 25 portfolios double sorted according to their size and value characteristics (see section 3.2), and another with a larger cross-section of 370 single-sorted decile portfolios from 37 characteristics, finding five economic meaningful factors that explain returns and achieve a Sharpe ratio twice as large as standard PCA. Recently, Bryzgalova et al. (2023) develop a framework that adds economically motivated moment targets to PCA, thereby nesting the proposal of Lettau and Pelger (2020a). In their empirical application, they work with the same 370 portfolios, but also consider 127 macroeconomic variables from McCracken and Ng (2016) to enforce non-zero correlation between the underlying factors and the fundamental shocks to the economic variables.

### 2.2.2 Bayesian procedures

An alternative approach is to replace parametric restrictions and instead impose either proper or diffuse Bayesian priors in the estimation of expected returns, variances and covariances (see Bawa, Brown and Klein (1979) for a review of the early literature and Avramov and Zhou (2010) for a more recent one, as well as Fabozzi, Huang and Zhou (2010) for the relationship between Bayesian and robust estimation methods). As an illustration, we focus on Frost and Savarino (1986), who assumed that the vector of prior means is

$$\dot{\boldsymbol{\mu}} = \dot{\mu}\boldsymbol{\iota}_n, \tag{8}$$

where $\boldsymbol{\iota}_n$ denotes a vector of $n$ ones, while the matrix of prior variances and covariances is given by

$$\dot{\boldsymbol{\Sigma}} = \dot{\sigma}^2 \left[\dot{\rho}\boldsymbol{\iota}_n\boldsymbol{\iota}_n' + (1-\dot{\rho})\mathbf{I}_n\right], \tag{9}$$

with $\dot{\sigma}^2 > 0$ and $1/(1-n) < \dot{\rho} < 1$. Note that the parameters $\dot{\mu}$, $\dot{\sigma}$ and $\dot{\rho}$ correspond to the common prior mean, standard deviation and correlation, respectively, of the assets under consideration. This restricted equicorrelated structure means that *a priori*, Frost and Savarino (1986) regard all assets as exchangeable, with the same expected excess returns and standard deviations, and equal pairwise correlations. Effectively, this means that before looking at the actual return data, their ignorance would be equally spread across assets.

The Sherman-Morrison-Woodbury formula (1) immediately implies that

$$\dot{\boldsymbol{\Sigma}}^{-1}\dot{\boldsymbol{\mu}} = \left[\frac{1}{1+(n-1)\dot{\rho}}\frac{\dot{\mu}}{\dot{\sigma}^2}\right]\boldsymbol{\iota}_n,$$

which means that the usual MV rule would lead to equal prior weights across assets regardless of the values of $\dot{\mu}$, $\dot{\sigma}$ and $\dot{\rho}$. Nevertheless, those parameters are important because the posterior values of the expected returns and covariance matrix will indeed depend on the values of $\dot{\mu}$, $\dot{\sigma}^2$ and $\dot{\rho}$. In particular, *a posteriori* the first two moments will be given by the expressions

$$\ddot{\boldsymbol{\mu}} = \omega_\tau\dot{\boldsymbol{\mu}} + (1-\omega_\tau)\bar{\boldsymbol{\mu}}$$

and

$$\begin{aligned}\ddot{\boldsymbol{\Sigma}} &= \frac{(\nu+T)}{(\nu+T-2)}\left(1+\frac{1}{\tau+T}\right) \\ &\quad\times\left[\omega_\nu\dot{\boldsymbol{\Sigma}} + (1-\omega_\nu)\bar{\boldsymbol{\Sigma}} + \omega_\tau\left(\frac{T}{T+\nu}\right)(\bar{\boldsymbol{\mu}}-\dot{\boldsymbol{\mu}})(\bar{\boldsymbol{\mu}}-\dot{\boldsymbol{\mu}})'\right],\end{aligned}$$

where $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Sigma}}$ are the vector of sample averages and covariance matrix of the excess returns, respectively, and

$$\begin{aligned}\omega_\tau &= \frac{\tau}{\tau+T} \\ \omega_\nu &= \frac{\nu}{\nu+T}\end{aligned}$$

capture the strength of the priors on means and variances, respectively, as determined by the parameters $\tau$ and $\nu$. In this sense, it is convenient to understand $\tau$ and $\nu$ as the sample lengths of hypothetical prior samples that have been used to come up with $\dot{\boldsymbol{\mu}}$ and $\dot{\boldsymbol{\Sigma}}$, respectively. Thus, $\tau = \nu = 0$ would simply equate the posterior values to the sample ones, while $\tau, \nu \to \infty$ would correspond to dogmatic priors, which give no weight to the sample information.

In practice, Frost and Savarino (1986) suggested an empirical Bayes procedure[4] with the following "estimators" of the prior hyperparameters:

$$\dot{\mu} = \frac{\boldsymbol{\iota}_n'\bar{\boldsymbol{\mu}}}{n}, \tag{10}$$

[4] Jorion (1986) also uses an empirical Bayesian approach that assumes (8) as prior means, but treats $\boldsymbol{\Sigma}$ as if it were known. In contrast, Ledoit and Wolf (2003) suggest an empirical Bayes estimator for $\boldsymbol{\Sigma}$ that combines the sample covariance matrix with Sharpe's (1963) market model ignoring the estimation of expected returns.

$$\dot{\sigma}^2 = \frac{tr\left[\bar{\boldsymbol{\Sigma}} + (\bar{\boldsymbol{\mu}} - \dot{\mu}\boldsymbol{\iota}_n)(\bar{\boldsymbol{\mu}} - \dot{\mu}\boldsymbol{\iota}_n)'\right]}{n} \tag{11}$$

and

$$\dot{\rho} = \frac{\boldsymbol{\iota}_n'\left[\bar{\boldsymbol{\Sigma}} + (\bar{\boldsymbol{\mu}} - \dot{\mu}\boldsymbol{\iota}_n)(\bar{\boldsymbol{\mu}} - \dot{\mu}\boldsymbol{\iota}_n)'\right]\boldsymbol{\iota}_n - tr\left[\bar{\boldsymbol{\Sigma}} + (\bar{\boldsymbol{\mu}} - \dot{\mu}\boldsymbol{\iota}_n)(\bar{\boldsymbol{\mu}} - \dot{\mu}\boldsymbol{\iota}_n)'\right]}{n(n-1)\dot{\sigma}^2}. \tag{12}$$

Consequently, Frost and Savarino (1986) shrink the MV weights towards the equally weighted portfolio, which is known to provide much better out-of-sample performance than the "plug-in" version that replaces the first and second moments in Markowitz's (1952) expressions by their sample counterparts, as shown by De Miguel, Garlappi and Uppal (2009) and Yuan and Zhou (2023), especially as the number of assets becomes large.

Alternative suggestions that ignore expected returns are the minimum variance portfolio $\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_n$, which Kan and Zhou (2007) combine with the usual plug-in MV portfolio to correct for the sampling uncertainty of the latter, and the risk parity portfolio, which effectively equalizes the marginal contribution of each asset to the standard deviation of the portfolio.[5] Interestingly, it is easy to show that the minimum variance portfolio and the risk parity one would coincide if the covariance matrix had an equicorrelated structure (see Maillard, Roncalli and Teïletche (2010)), and will further reduce to the equally weighted portfolio when (9) holds.

Other popular Bayesian-type procedures that impose structure on expected returns were suggested by Black and Littermann (1992), who use the market portfolio as an equilibrium anchor but allow for other prior views (see section 4 of Peñaranda (2008) for details), and Pástor (2000) and Pástor and Stambaugh (2000), whose prior means effectively impose the linear factor pricing model with traded factors in (5).

### 2.2.3 Shrinkage

Sometimes, it might be preferable to consider a procedure that effectively shrinks all the way to 0 the weights of many but not all assets. To see how this can be achieved, it is convenient to go back to the first exactly moment conditions in (2), in which $\boldsymbol{\phi}^\circ$ are the weights of the mean representing portfolio in the space of zero-cost portfolios. Trivially, the ideal mean representing portfolio would be 1 because $E(\mathbf{r}\cdot 1) = E(\mathbf{r})$. However, lack of arbitrage opportunities arguments imply that 1 cannot coincide with the payoffs of a zero-cost portfolio. Therefore, $\boldsymbol{\phi}^{\circ\prime}\mathbf{r}$ must coincide with the payoffs of the zero-cost portfolio that is closest to 1. As a consequence, $\boldsymbol{\phi}^\circ$ are the coefficients of the least squares projection of 1 onto the linear span of $\mathbf{r}$. This means that in practice, one can estimate the optimal weights by simply running the ordinary least squares (OLS) regression of $\boldsymbol{\iota}_T$, which is a vector of $T$ ones, onto $\mathbf{r}_t$ $(t = 1, 2, \ldots, T)$.

[5] The marginal risk contribution of an asset is proportional to the so-called incremental value at risk (iVaR) under the assumption that the joint distribution of returns is Gaussian with a zero mean.

The problem is that the elements of $\mathbf{r}$ will be very highly collinear when $n$ is large, which implies that the estimated regression coefficients may be numerically unreliable. Although collinearity should not affect much the properties of the optimal portfolio, it certainly distorts the interpretation of the portfolio weights. Moreover, OLS breaks downs altogether when $n \geq T + 1$. For that reason, some authors suggest to substantially reduce the number of assets that are effectively selected. This can be achieved in different ways, but a popular choice is to impose restrictions on the weights, such as $0 \leq w_i \leq 1$ (see Jagannathan and Ma (2003) for a reinterpretation of certain constraints on portfolio weights as a specific shrinkage estimation of the covariance matrix $\mathbf{\Sigma}$). While the non-negative bounds may make sense if one does not want to short a specific asset, and can be easily imposed in the regression of $\boldsymbol{\iota}_T$ on $\mathbf{r}_t$, the upper bound is somewhat arbitrary given that the scaling of each individual payoff $r_i$ is to some extent arbitrary too.[6]

Let us consider an alternative approach that replaces the individual inequality constraints $0 \leq w_i \leq 1$ by an aggregate constraint of the form

$$||\mathbf{w}||_1 = \sum\nolimits_{i=1}^{n} |w_i| \leq \lambda.$$

Given that $\sum_{i=1}^{n} |w_i| = \sum_{i \in \ell} w_i + \sum_{i \in s} |w_i|$, where $\ell$ is the set of assets with positive weights and $s$ is the complementary set of assets with negative weights, we can interpret the above restriction as imposing an upper bound on the sum of short selling positions plus the sum of long positions. A simple procedure for achieving this goal would minimize the residual sum of squares in the regression of 1 on $\mathbf{r}$ subject to the above inequality constraint. More formally, we wish to find the solution to the program

$$\min_{\mathbf{w} \in \mathbb{R}^n} ||1 - \mathbf{r}'\mathbf{w}||_2^2 + \varrho ||\mathbf{w}||_1,$$

where $||.||_2$ denotes the usual Euclidean norm, and $\varrho$ is the Kuhn-Tucker multiplier associated to the inequality constraint above. The solution to this problem is known as the least absolute shrinkage and selection operator (lasso) regression. This procedure was first introduced by Tibshirani (1996), and it is increasingly becoming the default procedure to conduct stepwise regression and related data mining procedures.

Ao, Li and Zheng (2019) apply a slight variation of the lasso approach described above to construct optimal MV portfolios from the individual constituents of the Dow Jones Industrial

[6] Importantly, note that the properties of OLS regressions trivially imply that the payoffs to the optimally chosen portfolio will be invariant to the degree of leverage of the different elements of $\mathbf{r}$. Specifically, if we double the leverage of $r_i$ say, then the corresponding weight will get halved, but the fitted value will remain the same. Also note that the mean return associated to the fitted values from this regression will be $\boldsymbol{\mu}'\mathbf{\Gamma}^{-1}\boldsymbol{\mu}$, so if one would like to obtain a different expected return, say $\mu$, one should either proportionally scale up or down all the weights by $\mu/\boldsymbol{\mu}'\mathbf{\Gamma}^{-1}\boldsymbol{\mu}$, or the constant regressand by the same amount.

Average 30 and the S&P 500. The main difference of their approach is that they multiply the vector of ones that acts as regressand by either the constant $(1 + \dddot{\theta})\dddot{\theta}^{-1}\mu$ if the objective is to achieve a target expected excess return level $\mu$, or by $(1 + \dddot{\theta})\dddot{\theta}^{-1/2}\sigma$ when the objective is to achieve a standard deviation $\sigma$, where $\dddot{\theta}$ is the unbiased estimator of the maximum square Sharpe ratio in (3). By the numerical properties of OLS estimators, though, this re-scaling does not affect either the assets that end up with a non-negative weight or the relative values of those weights (see footnote 6).

The main advantage of expressing the inequality constraint in terms of the so-called $L_1$ (or absolute value) norm is that it leads to parsimonious solutions in which many of the optimal weights will be 0, effectively reducing the number of elements of $\mathbf{r}$ that will appear in the optimal portfolio. The parameter $\lambda$ plays a crucial role in the solution too, and therefore has to be chosen judiciously, since a large value of $\lambda$ will make the problem effectively unconstrained. In this regard, it is useful to interpret the lasso output as the posterior mode estimates obtained by a Bayesian regression procedure in which the prior views on the coefficient $w_i$ conform to a Laplace distribution with 0 mean and are independent across assets. The Laplace, or double exponential distribution, is a symmetric distribution characterized by both a high peak at the mode and fat tails. Apart from the corresponding differences in the kurtosis coefficient, which is 6 for the Laplace versus 3 for the normal, the main difference between Laplace and Gaussian distributions is that the density function of the former is continuous but not differentiable at 0. As a result, the mode of posterior distribution of $w_i$ will often be 0 too, which partly explains the large number of zero $w_i's$ obtained in the lasso solution.

In this context, the value of $\lambda$, or equivalently, the value of $\varrho$, can be understood as the parameter governing the strength of the priors. The bigger $\lambda$ is (or the smaller $\varrho$ is), the weaker the priors are, so that the solution will resemble more and more the unrestricted MV solution. In contrast, the smaller $\lambda$ is (or the larger $\varrho$ is), the stronger the prior, in which case the solution will differ substantially from the standard one. The relationship between the Laplace distribution and the $\| \, . \, \|_1$ norm is not surprising because the maximum likelihood estimator (MLE) of the location parameter of a Laplace distribution is the sample median, as opposed to the sample mean, which is the MLE under normality.

In the regression context that we are considering, an independent zero-mean Gaussian prior on the "regression coefficient" $w_i$ will be equivalent to replacing the penalty $||\mathbf{w}||_1$ by $||\mathbf{w}||_2$. This will shrink the unrestricted MV weights toward 0, but will not exactly set to 0 any of those weights, so that all individual subsystems in $\mathbf{r}$ will be assigned some weight. This is the so-called ridge procedure, which is commonly used to deal with multicollinearity. Interestingly,

both lasso and ridge procedures work when $n$ is bigger than $T$. For that reason, they are often referred to as regularization procedures.

Unlike ridge procedures, which have as closed-form solution for the weights the sample version of $(\mathbf{\Gamma} + \varrho \mathbf{I}_n)^{-1}\boldsymbol{\mu}$, lasso has to be computed numerically. A brute force approach to the solution would involve the computation of the optimal MV solution for every possible subset of the $n$ assets. Having obtained the optimal unrestricted weights for a particular set, one would then evaluate the objective function at those weights, and select the set for which this objective reaches the *minimum minimorum*. The problem with this approach is that there are $2^n$ possible subsets, which makes it infeasible for moderately large $n$. Fortunately, the lasso problem can also be solved using quadratic programming or more general convex optimization methods, as well as by means of specific algorithms. In particular, Efron et al. (2004) proposed a slight modification of the so-called Least Angle Regression (LARS) algorithm as a very fast way of computing all possible lasso regressions as a function of $\lambda$. An important advantage of this algorithm is that it will progressively incorporate variables to the regression as $\lambda$ increases by exploiting the fact that marginal increases in this parameter will typically lead to no change in the components of the optimal portfolio. An additional advantage already mentioned is that it will work even in contexts in which $n$ is larger than the number of observations.

In fact, it is possible to take a weighted combination of the $||\mathbf{w}||_1$ and $||\mathbf{w}||_2$ penalties, as in the so-called elastic net procedure, which tends to select strongly correlated returns together. Alternatives that impose a hierarchical structure by either including always some assets or grouping them so that they can only appear in combination with others, as in group lasso, offer attractive possibilities in some circumstances.

# 3 Conditioning information

So far, we have considered a portfolio problem in which the information agents may have at the time they make investment decisions plays no role. However, except for passive investors who merely track a broad stock market index such as the S&P 500, the Russell 3000 or one of its international counterparts, most asset managers systematically rely on information in making their choices. In this respect, there are two broad categories of active investors: stock pickers, who rely on the individual characteristics of the assets, and market timers, who try to predict trends in the prices of either some specific assets or the market as a whole. In practice, of course, most asset managers employ a combination of stock picking and market timing strategies.

Interestingly, before the stock market crash of October 1987, market timing was regarded with suspicion, but nowadays everybody agrees that the distribution of asset returns changes

frequently in predictable ways, if not necessarily in terms of their means (see Spiegel (2008)), at least in terms of their variances and correlations, and that investors can exploit this fact to their advantage by using conditional distributions in designing their portfolio strategies. In contrast, stock picking has recently lost its allure, especially in view of the fact that many active fund managers do not generate enough value added to compensate for their higher fees, with some even accused of being hardly distinguishable from index trackers.

We review the two strands of the literature in section 3.2 below, but, before doing so, it is convenient to discuss in some detail the subtle but important differences that the presence of conditioning information introduces in the theory of asset management in section 2.1, and its relationship to asset pricing.

### 3.1 Theoretical background

For simplicity, suppose that there are $d$ variables, $\mathbf{x} = (x_1, \ldots, x_d)'$, sometimes called instruments or signals, which help predict asset returns. These predictors could be individual asset characteristics, such as the size of a firm in terms of its market capitalization or its book to market ratio, or macroeconomic time series, such as the inflation level or the spread between corporate and government bonds. We also assume that there are only three important dates in this economy: the decision, trading, and payoff dates. Investors design ex ante portfolio strategies at the decision date which may depend on the information that they will observe at the trading date. Finally, they receive payoffs at the final date. In a world with multiple periods, this approach leads to a myopic portfolio decision, but we postpone the discussion of intertemporal portfolio choice until section 4.1.

We denote the set of all random variables that are measurable with respect to $\mathbf{x}$ by $I$. In this context, we denote the first two conditional moments of the available zero-cost payoffs by $E(\mathbf{r}|I)$ and $E(\mathbf{r}\mathbf{r}'|I)$, respectively. We treat these moments as random variables that belong to $I$, whose realizations correspond to the values they take for a specific value of $\mathbf{x}$. To avoid a trivial uninformative set up, we assume that not all these random variables are degenerate. We also assume that the diagonal elements of $E(\mathbf{r}\mathbf{r}'|I)$ are uniformly bounded with probability one, so that a fortiori all the elements of $\mathbf{r}$ belong the collection of random variables defined on the underlying probability space with bounded unconditional second moments. Regarding the conditional covariance matrix of $\mathbf{r}$, $Var(\mathbf{r}|I)$, we assume its smallest eigenvalue is uniformly bounded away from 0 with probability one, which implies that none of the zero-cost portfolios in $\mathbf{r}$ is either conditionally riskless or redundant, and moreover, that it is not possible to generate a conditionally riskless portfolio from $\mathbf{r}$ other than the trivial one. In what follows, we will refer to the conditional span of $\mathbf{r}$ as the payoff space $P$. In this context, $\mathbf{w} \in I$ indicates an *active*

portfolio strategy, while a vector of constant weights $\mathbf{w} \in \mathbb{R}^n$ indicates a *passive* portfolio.

We can then define the uncentred conditional mean representing portfolio in the space of zero-cost portfolios as $r^\circ$, so that $E(\mathbf{r}|I) = E(\mathbf{r}r^\circ|I)$. By analogy to the case without conditioning information in section 2.1,

$$r^\circ = \mathbf{r}'E^{-1}(\mathbf{r}\mathbf{r}'|I)E(\mathbf{r}|I), \tag{13}$$

which is the *excess* return that "mimics" a safe asset with constant payoffs 1 with the minimum "conditional tracking error". Then, if we rule out risk neutral pricing, so that not all conditional expected returns are equal, it is possible to prove that all the portfolios on the *zero-cost* conditional MV frontier generated from *all* primitive assets will be spanned by $r^\circ$. Therefore, this portfolio achieves the maximum conditional Sharpe ratio among all arbitrage portfolios.

From the practical point of view, a problem with this conditional MV frontier is that it depends on the information set an investor has at her disposal. In contrast, outsiders usually only observe the excess returns to her portfolio. Specifically, the composition of the portfolios held by professional managers is either a snap-shot taken only once per quarter, as in the mutual fund industry, or not at all, as in the case of hedge funds and other alternative investment managers such as commodity trading advisors. Therefore, given that performance evaluation must often be based exclusively on observed track records, it still makes sense to study MV frontiers based on unconditional moments even though investors are using conditional ones. In this respect, Hansen and Richard (1987) define the unconditional MV frontier for zero cost portfolios as the *highest* lower bound on the unconditional variance for each level of expected return that can be achieved by portfolios with weights that may depend on conditioning information, but whose cost is 0. Thus, the unconditional arbitrage MV frontier will be given by the set of active portfolio strategies that solve the problem

$$\min_{p \in P} E(p^2) \qquad s.t. \qquad E(p) = \nu \in \mathbb{R}, \qquad C(p|I) = 0. \tag{14}$$

Hansen and Richard (1987) show that the excess returns that solve (14) correspond to the following passive portfolio of $r^\circ$:

$$p_U(\nu) = \omega_U(\nu)r^\circ, \tag{15}$$

where the constant $\omega_U(\nu)$ guarantees that the constraint $E[p_U(\nu)] = \nu$ is satisfied. This frontier is related to the unconditional SDF frontier introduced by Gallant, Hansen and Tauchen (1990), which yields the highest lower bound on the variance of SDFs that correctly price any portfolio whose weights may also depend on conditioning information. Unlike what happens in the case of assets of arbitrary cost, these two frontiers are dual to each other for zero-cost portfolios (see Peñaranda and Sentana (2016) for further details).

Importantly, the set of unconditionally MV efficient portfolios is a subset of the set of conditionally efficient ones characterized by a constant weight on the conditional mean representing portfolio $r^\circ$. This seemingly cryptic remark has important consequences for portfolio managers. For example, the trading rule implicit in (13) is different from the natural counterpart to the static MV rule that we have discussed in section 2.1, whose weights would be $E'(\mathbf{r}|I)V^{-1}(\mathbf{r}|I)$. In particular, the Sherman-Morrison-Woodbury formula (1) applied to the conditional second moments implies that

$$E^{-1}(\mathbf{r}\mathbf{r}'|I)E(\mathbf{r}|I) = \left[\frac{1}{1+E'(\mathbf{r}|I)V^{-1}(\mathbf{r}|I)E(\mathbf{r}|I)}\right] V^{-1}(\mathbf{r}|I)E(\mathbf{r}|I), \tag{16}$$

which means that although the relative weights are the same, the factor of proportionality varies with the information. The difference can perhaps be seen more clearly in the single asset case, in which

$$r^\circ = \frac{E(r|I)}{V(r|I)+E^2(r|I)}r = \left[\frac{1}{1+E^2(r|I)/V(r|I)}\right]\frac{E(r|I)}{V(r|I)}r.$$

As a result, while the conditional Sharpe ratio is the same for both strategies, the unconditional Sharpe ratio of the trading rule based on second moments (13) is always at least as large as the unconditional Sharpe ratio of the actively trading rule based on conditional variances and covariances, with equality if and only if the maximum conditional Sharpe ratio $E'(\mathbf{r}|I)V^{-1}(\mathbf{r}|I)E(\mathbf{r}|I)$ is constant (see Peñaranda (2016) for further details).[7] In fact, Sentana (2005) provides counterexamples in which the unconditional Sharpe ratio of this last strategy could be lower than the Sharpe ratio of an investment fund that follows a simple buy and hold strategy for the underlying asset. Sentana (2005) also shows that unconditional MV applied to the payoff vector $(1, \mathbf{x}') \otimes \mathbf{r}'$ results in a portfolio that maximizes the unconditional Sharpe ratio among all portfolios of $\mathbf{r}$ whose weights are affine functions of $\mathbf{x}$, like the ones we analyze in section 3.2.2 below.

Managed portfolios whose leverage depends linearly on $\mathbf{x}$ are often regarded as an ad-hoc way of approximating the effect of conditioning information. This criticism is less worrying than it may seem at first sight because $\mathbf{x}$ can contain non-linear transformations of a relatively small set of observed variables. In fact, Peñaranda and Sentana (2016) and Chen et al. (2023) show that suitably selected managed portfolios can approximately replicate the payoff space $P$ with solid statistical foundations. Specifically, these papers explicitly relate managed portfolios to sieve methods in a new econometric methodology called "sieve managed portfolios". As is well known, sieves effectively approximate the functions to be estimated, the optimal active portfolio

[7] It is worth mentioning that there are other relevant subsets of conditionally efficient returns apart from unconditionally efficient returns. Specifically, Peñaranda and Wu (2022) study the subset of conditionally efficient portfolios that achieve a constant conditional expected return (a return target), and the subset that achieve a constant conditional variance (a risk target). They show that risk targeting generates better unconditional performance than return targeting across a wide range of metrics.

weights in this case, by means of parameter spaces whose dimension increases with sample size (see Chen (2007) for a survey of sieve methods in econometrics). Let $\mathbf{b}^{d_T}(\mathbf{x})$ denote a $d_T \times 1$ vector of known sieve basis functions (power series, splines, Fourier series, etc.) with the property that its linear combinations can approximate arbitrarily well any square integrable real-value function of $\mathbf{x}$ by allowing the so-called mesh size parameter $d_T$ (number of terms in a series or number of knots in a spline) to increase without bound. Since the approximating spaces are characterized by a finite number of parameters, sieve methods effectively reduce the estimation problem to a parametric one. Nevertheless, at some point the number of underlying strategies will become very large, and some regularization would become necessary. This regularization can be easily achieved in the context of the regression of 1 on the excess returns of the sieves managed portfolios, $\mathbf{b}^{d_T}(\mathbf{x}) \otimes \mathbf{r}$, by means of the lasso or ridge procedures described at the end of section 2.2.3.

### 3.2 Big data on asset characteristics and macroeconomic indicators

According to the CAPM, the market portfolio is MV efficient, and consequently, investors should be able to attain the maximum Sharpe ratio by simply buying an exchange traded fund (ETF) that replicates a broad stock market index such as the S&P 500, the Russell 3000, or their international counterparts, such as the MSCI ACWI. The applied finance literature of the early seventies broadly supported the view that the stock market as a whole was efficient, and that the static CAPM provided a valid representation of the relative valuation of many assets. In contrast, the late seventies and early eighties upended the status quo. Specifically, many studies found that several asset-specific characteristics that in principle should play no role were useful in explaining the difference between actual risk premia and the theoretical risk premia implied by the CAPM. The first examples were size (see Banz (1981)) and book-to-market ratios, which allow the classification of firms into mutually exclusive value and growth groups (see Statman (1980) and Rosenberg, Reid and Lanstein (1985)). These two examples were soon followed by others combining accounting and market price data, or only the latter (see De Bondt and Thaler (1984) for reversals and Jegadeesh and Titman (1993) for momentum). In fact, Asness, Moskowitz and Pedersen (2013) mention that professional portfolio managers have long considered strategies that exploit individual stock characteristics such as value and momentum.

As usual, these anomalies could be interpreted as either a failure of market rationality or the asset pricing model used. Therefore, it is not surprising that new empirically oriented asset pricing models came to the fore. One of the most influential models was Fama and French (1993) 3-factor model in which the affine SDF included not only the excess returns on the market

portfolio, but also the excess returns of two other trading strategies that aim to capture the size and value effects: SMB, which means long/short in small/large capitalization stocks, based on said size effect, and HML, or long/short in high/low book-to market ones, which exploits the fact that the returns to value stocks tend to outperform those to growth stocks in the long run.

Typically, factor or style portfolios are calculated as follows. At a given point in time, all individual stocks in the CRSP database are ranked according to a characteristic obtained either from past returns, as in the case of momentum and short- and long-term reversals, or by combining price data with variables in the income statements and balance sheets of publicly traded companies reported in the Compustat database. Then, a zero-cost portfolio is created by taking long positions in those stocks for which the value of their characteristic exceeds the 100-x% percentile and short positions in those with values below the x% percentile, with x equal to 10 for example. In the case of momentum and reversal factors, the rebalancing takes place monthly, while for characteristics based on accounting variables, it takes place once a year in early July after most companies have officially filed their annual accounts. Sometimes, though, these univariate sorts are replaced by bivariate and even trivariate ones. In particular, the original size and value factors are based on six portfolios computed from the intersections of two portfolios formed on size with a breakpoint equal to the median NYSE market equity at the end of June, and three portfolios formed on the ratio of book equity to market equity at the end of the previous year with breakpoints the 30th and 70th NYSE percentiles.

After the success of the Fama and French (1993) model, additional pricing factors were proposed. For example, Carhart (1997) considered a fourth factor to capture "momentum", which goes long in the winners over the previous 12 months (excluding the most recent one) and short in the corresponding losers. Given the interest of financial market participants and finance journals in these empirical results, researchers progressively entertained a broader and broader set of factors, which has resulted in several success claims. Harvey, Liu and Zhu (2016) contained a comprehensive list of references theretofore, which included 315(!) different factors, and the so-called "factor zoo" has only increased since, raising concerns that many empirical asset pricing models are overspecified (see Manresa, Peñaranda and Sentana (2023)). To ease replicability, Chen and Zimmermann (2022) have created a repository of many of these characteristic-sorted portfolios.

### 3.2.1 Forecasting returns

The most common way of incorporating conditioning information in portfolio decisions is to use it for the purposes of predicting returns. As is well known, forecasting is one of the most frequent examples of supervised learning in the machine learning literature, which can

accommodate a large number of characteristics, macroeconomic variables and their interactions. The estimated forecasting function is then used to build portfolios, typically by means of long-short strategies obtained from decile spreads.

A very popular forecasting method is regression trees, which extends classification trees to situations in which the dependent variable is continuous rather than binary. Tree-based methods have two main features. First, they approximate the conditional mean functions by means of step functions defined over a partition of the space of regressors into hyperrectangles or splits. Second, given that it is generally not possible to obtain the globally best partition in a reasonable amount of time, these methods are implemented by means of a greedy algorithm that finds a sequence of optimal splits for a single variable at a time. A single tree is easy to interpret but it suffers from high sampling variability. Therefore, in practice these procedures are implemented through ensemble methods that average many trees, even though this approach loses the simple interpretation of a single tree. One common variant is bagging, where trees are trained with different bootstrapped samples run in parallel. Another common variant is boosting, where a sequence of trees are trained on modified versions of the data that depend on the previous model errors.

For example, Moritz and Zimmermann (2016) use random forests to obtain the conditional mean of stock returns given an information set of past returns and firm characteristics. Random forests are an example of bagging, but allowing for a random set of predictors at each split. These authors find that a trading strategy that goes long in stocks with high return predictions and short in stocks with low return predictions achieves an information ratio twice as high as Fama-MacBeth regressions that allow for interactions between the characteristics. As is well known, the information ratio is analogous to the Sharpe ratio but for the returns of an asset relative to its benchmark. In turn, Fama and MacBeth (1973) regressions are a popular procedure of rolling cross-sectional regressions that generates a time-series of intercept and slope coefficients whose values can be related in some applications to the prices of risk that appear in the SDF. In each cross-sectional regression, the dependent variable is the next-period return for each stock, and the explanatory variables the current stock characteristics.

More recently, Gu, Kelly and Xiu (2020) apply a suite of machine learning methods to 920 covariates obtained by combining 74 industry dummies with the interactions of 94 characteristics and 8 macroeconomic variables. In particular, they consider OLS, dimensionality reduction with principal components regression and partial least squares, regularization with elastic net and group lasso, random forests and gradient boosted trees, as well as several neural networks (NNs), finding that decile spread portfolios constructed using NNs dominate the other proce-

dures in terms of Sharpe ratios and peak to trough price falls or "drawdowns". A simple NN with a single hidden layer computes several linear combinations of the predictors and applies a nonlinear transformation to each one of them. The output of the NN, a return forecast in our setting, is a linear combination of those nonlinear transformations. One can also compute a more complex NN with several hidden layers by compounding the linear combinations and their nonlinear transformations, the archetypal example of deep learning. NNs are usually estimated or "trained" by an algorithm called stochastic gradient descent, which in this case applies the chain rule across the layers in a procedure known as "backpropagation". In this section, we consider fully connected feedforward networks, describing more recent architectures in subsequent ones.

On their part, Freyberger, Neuhierl and Weber (2020) use an additive nonparametric approach with no interaction across characteristics, in which each component is approximated by quadratic splines. They also use group lasso to select characteristics, penalizing jointly all the spline coefficients associated to each one of them. Although they work with 62 characteristics, they find that less than twenty provide incremental information, with their nonparametric model more than doubling the Sharpe ratio of a standard linear one.

Using the characteristics data of Freyberger, Neuhierl and Weber (2020), Kelly, Pruitt and Su (2019) and Kim, Korajczyk and Neuhierl (2021) develop latent factor models with time-varying intercepts and loadings that depend on those characteristics. Those models can be used to construct arbitrage portfolios that eliminate factor risk.

In turn, Chen and Zimmermann (2022) work with 319 characteristics, finding that their predictive ability is robust and credible. In contrast, Chen and Velikov (2023), who focus on the profitability of long-short portfolios instead, do not find any profitable rule after taking into account trading costs and the staleness of some time series.

On the other hand, Han et al. (2021) develop a cross-sectional approach based on Fama-MacBeth regressions that uses lasso, forecast combination and forecast encompassing. They work with 193 characteristics from Chen and Zimmermann (2022), and construct a portfolio strategy with long and short positions that yields a high Sharpe ratio and an abnormal expected return or "alpha". They also confirm that many characteristics are able to forecast cross-sectional differences in expected returns.

The previous papers focus on US stocks but others provide international evidence. Specifically, Rasekhschaffe and Jones (2019) use several machine learning methods for stock selection with 194 characteristics. They study both the US and a portfolio of several other developed markets. The methods that they consider are a gradient boosted regression tree, a NN, and some

bagging estimators, illustrating the benefits of their methods by means of a long-short strategy of decile spreads. In addition, Tobek and Hronec (2021) consider stocks from 23 developed countries and 153 characteristics. They implement several machine learning methods along the lines of Gu, Kelly and Xiu (2020). They conduct their analysis using liquid stocks for the US, Japan, the Asia-Pacific region, and Europe, finding a significant profitability associated to their methods, with the main source of information for all the geographic areas being the US. More recently, Choi, Jiang and Zhang (2024) use stocks from 32 international markets, finding that the use of 36 characteristics for US firms using NNs improves the return predictions in other countries.

Finally, Leipold, Wang and Zhou (2022) also apply the methods of Gu, Kelly and Xiu (2020) to thousands of Chinese stocks using 94 characteristics, 80 industry dummies, and 11 macroeconomic variables as predictors. They find some distinctive features relative to the US market, such as a more relevant role for liquidity and a lesser one for momentum. However, they are forced to restrict their analysis to the performance of long-only portfolios, which remains economically significant, because shorting is not really practical in the Chinese stock market.

All previous papers forecast returns, but another possibility is to forecast corporate earnings. In this respect, Cao and You (2024) find that machine learning methods, especially nonlinear ones, generate significantly better forecasts than analysts' consensus forecasts. They consider, in particular, lasso and ridge regressions, random forests, boosted trees and NNs with 60 accounting predictors at the annual frequency, finding that the corresponding portfolios based on quintile spreads yield statistically significant average returns.

A different strand of the literature studies monthly market timing procedures using the 17 Goyal and Welch (2008) macroeconomic predictors.[8] In particular, Rossi (2018) uses boosted trees to predict stock market returns and their volatilities. His volatility forecast also makes use of 250 past squared daily returns using a mixed data sampling (MIDAS) model. He studies the performance of an MV strategy that uses those forecasts and another one that also uses boosted trees but this time to forecast the optimal MV weight directly, finding that both approaches translate into profitable strategies. In turn, Jacobsen, Jiang and Zhang (2019) develop a mixture of bagging and boosting to generate average return forecasts using methods such as lasso, Bayesian model averaging, and weighted-average least squares. Then, they combine these return forecasts with a variance forecast obtained from a five-year rolling window to compute the MV weight, which they use next in their market timing strategies, finding that the gains

[8] More recently, Goyal, Welch and Zafirov (2023) add 29 predictors to the original Goyal and Welch (2008) variables, finding that only a few of them forecast reasonably well the equity premium for the market both in-sample and out-of-sample.

from their method are mainly concentrated in periods of extreme market falls.

In an example of assets different from stocks, Bianchi, Büchner and Tamoni (2021) apply machine learning to US Treasury bonds. They carry out two empirical applications, one that forecasts bond returns using the cross-section of yields, and another one that uses both forward rates and 128 macroeconomic variables described in McCracken and Ng (2016). Bianchi, Büchner and Tamoni (2021) find that both extreme trees, which use random partitions, and NNs are able to achieve a significant bond return predictability. The portfolio choice criteria that they study include an MV investor and another one with constant relative risk aversion (CRRA) preferences. In the MV weight formula, in particular, they combine machine learning methods for the mean with a rolling estimator for the variance, observing that their forecasts yield large economic gains. Similarly, Kelly, Palhares and Pruitt (2023) apply the methodology of Kelly, Pruitt and Su (2019) to corporate bonds. They work with 14,600 bonds and 29 bond/firm characteristics, finding that their factor model improves earlier corporate bond models proposed in the literature by a large margin. In addition, they find that the credit risk premium is notably larger than previously estimated, and that there is a closer integration between debt and equity markets than usually thought.

### 3.2.2 Optimal weights

In this section, we review several research papers that compute portfolio weights directly by optimizing some economic criterion without the need to forecast returns. Nevertheless, most of them still consider monthly portfolio weight rebalancing for thousands of US stocks. Brandt, Santa-Clara and Valkanov (2009) handle a large cross-section of $N_t$ assets at the monthly frequency with a few characteristics by using the following parametrization of the portfolio weights:

$$w_{it} = \bar{w}_{it} + \frac{1}{N_t}\boldsymbol{\theta}'\hat{\mathbf{x}}_{it}, \tag{17}$$

where $\bar{w}_{it}$ represents the weight on asset $i$ of a benchmark portfolio, typically a value-weighted index, and $\hat{\mathbf{x}}_{it}$ vector the cross-sectionally standardized characteristics for this asset. They then estimate the vector of coefficients $\boldsymbol{\theta}$ so as to maximize the in-sample value of an evaluation criterion such as the Sharpe ratio. Previously, Brandt and Santa-Clara (2006) used parametric weights in the context of asset allocation across time on the basis of macroeconomic predictors. Specifically, they studied the performance of market-timing strategies for stocks, bonds and cash in which they allowed the weights of each asset class to be a different affine function of the common vector of predictors.

Several recent papers build on (17) to combine many characteristics. For example, Kozak,

Nagel and Santosh (2020) translate such a strategy to an asset pricing context. Parametrizing the SDF weights as affine functions of the characteristics allows them to work with 80 characteristic-based portfolios instead of the original large cross-section of individual stocks. When they also consider pairwise interactions of individual characteristics, the number of managed portfolios raises to 3,400. In addition, they include a ridge or lasso regression penalization in estimating the SDF. Given that the SDF weights have an optimal MV portfolio interpretation, as we have seen in previous sections, these authors can generate alpha relative to the Fama and French (2015) model augmented with a momentum factor. Although their results might seem as evidence against six factors being sufficient to summarize the cross-section of stocks, they also show that five principal components of the managed portfolios render the alpha of the optimal MV portfolio insignificant.

Another paper that also builds on Brandt, Santa-Clara and Valkanov (2009) is De Miguel et al. (2020), who show that transaction costs increase the number of characteristics that are jointly significant for optimal portfolios. They work with MV parametric portfolios that are affine functions of 51 characteristics. Apart from explicitly considering transaction costs, they use lasso regularization, finding that transaction costs increase the number of significant characteristics from 6 to 15. Kelly, Malamud and Pedersen (2023) also maintain that the portfolio weights are affine in the characteristics, as in (17), but allow the characteristics of other assets to affect the weights of asset $i$, thereby allowing their portfolios to exploit cross-predictability across assets.

More recently, Simon, Weibels and Zimmermann (2023) substantially depart from the original spirit of parametric strategies by maximizing expected utility when the portfolio weights, as functions of the characteristics, are modelled using deep NNs that allow for interactions between characteristics. Using 157 characteristics from Chen and Zimmermann (2022), they find significant utility gains from their deep learning procedures. Interestingly, they also find that risk aversion can be interpreted as a regularization parameter.

Bryzgalova, Pelger and Zhu (2023) also abandon the parametric framework of Kozak, Nagel and Santosh (2020) and use decision trees to build cross-sections of stocks based on characteristics. Their use of trees, though, is very different from the usual greedy algorithm in machine learning. For a given set of characteristics, at each step they group stocks with a tree obtained by splits based on median characteristics, limiting the depth to four splits to obtain well-diversified managed portfolios. They consider all possible trees derived by different orderings of the characteristics. Then they prune those trees using the following three tuning parameters: shrinking means and variances, and a lasso penalty for the SDF coefficients, which they select by maximizing in the validation sample the Sharpe ratio of the optimal MV portfolio linked to the SDF.

Somewhat surprisingly, they show that their SDF framework is equivalent to an MV framework with an elastic net penalty. In their empirical application, they initially consider triple-sorted portfolios but finally settle for analyzing 10 characteristics jointly. An important result is that a mere 20 to 40 tree-based portfolios can double the maximum Sharpe ratio associated to the underlying 100 decile-sorted, long/short portfolios.

Cong et al. (2023) also use decision trees to construct the MV frontier from an unbalanced panel of stocks. They rely on global split criteria based on maximum Sharpe ratios, growing a tree until 10 basis portfolios are generated, and considering a sequence of boosted trees. A managed portfolio is then constructed for each one of those trees, and the final output is the MV frontier for the corresponding cross-section of managed portfolios. They work with 61 characteristics jointly with 10 macro variables, and find a significant improvement of the MV frontier thus obtained with respect to more standard portfolio construction techniques.

On their part, Goulet Coulombe and Göbel (2023) use the Goyal and Welch (2008) macroeconomic predictors ignoring characteristics data. They develop an algorithm that optimizes portfolio weights so that the portfolio is maximally predictable in the sense of Lo and MacKinlay (1997). The algorithm iterates between a random forest step for the portfolio return prediction and a ridge regression step for the portfolio weights, finding a significant increase in profitability despite using relatively little conditioning information.

In a very recent paper, Chen, Pelger and Zhu (2024) develop an alternative approach that differs from the existing literature in several important aspects. They combine 46 characteristics and 178 macroeconomic predictors, most of which they obtain from McCracken and Ng (2016). Next, they estimate the SDF implied by the cross-section of stocks using a minimax criterion. Specifically, they separate the managed portfolios that appear in the SDF from the managed portfolios that must be priced, minimizing the pricing errors of the former, but maximizing them for the latter. They parametrize both sets of managed portfolios by a feedforward network that combines stock characteristics with macroeconomic regimes obtained by means of a long short-term memory network (LSTM), which we describe in some detail in section 4.2 in the context of textual data. They train their model as a generative adversarial network (GAN),[9] with several hyperparameters that they obtain by maximizing over the validation data the Sharpe ratio associated to the SDF. They show that their method outperforms several benchmarks in terms of this portfolio performance criterion.

Finally, Avramov, Cheng and Metzker (2023) study the economic significance of several

[9] The GAN architecture is based on two networks with opposite goals. It was developed in the area of computer vision to generate images close to real ones, with the two competing networks being a generator that creates images from noise, and a discriminator that tries to distinguish those fake pictures from real ones.

machine learning methods. They consider the two portfolio construction approaches we have already discussed, namely decile spread portfolios obtained from forecasts, as Gu, Kelly and Xiu (2020), whose covariates they share, and optimal weights, like Chen, Pelger and Zhu (2024). Imposing economic restrictions, they find that the profitability of deep learning methods becomes considerably lower when they exclude microcaps, distressed stocks, or episodes of high market volatility. They also report that the portfolio turnover of machine learning methods is considerably higher than the turnover associated to traditional portfolios based on individual characteristics, which means that transaction costs cannot be ignored in practice. Another of their findings is that deep learning strategies are considerably more profitable during periods in which investor sentiment and volatility are high and market liquidity low. Finally, they report that machine learning methods are economically interpretable, in the sense that they identify stocks that are mispriced according to well-established empirical regularities, and they display less downside risk too.

### 3.3 Interpretability and complexity

Some machine learning methods are often considered black boxes that hide the connection between inputs and outputs. Nevertheless, some of the papers mentioned in previous sections go to great lengths to explicitly address this concern. For example, Moritz and Zimmermann (2016) compute a measure of predictor importance based on the change in mean square error (MSE) when a predictor is randomly permuted. They also look at the partial derivatives of the MSE with respect to each predictor, specifically checking if the model predictions can be explained by a simple linear regression. Similarly, Simon, Weibels and Zimmermann (2023) consider variable importance, partial dependence, and two surrogate models: a linear one and another one with interactions. In turn, Gu, Kelly and Xiu (2020) compute the reduction in the predictive $R^2$ of their panel of stocks when all values of a predictor are set to zero, as well as the sum of squared partial derivatives of the MSE with respect to each input variable. They find that the most powerful predictors are variations of momentum, liquidity, and volatility. Lastly, Chen, Pelger and Zhu (2024) rank variables by their sensitivity, which they measure by the average absolute derivative of the SDF weights with respect to each variable.

Li, Turkington and Yazdani (2020) also emphasize that machine learning methods for investment must be interpretable. For that reason, they develop a method built on the concept of partial dependence in Friedman (2001), which isolates the linear and nonlinear effects of each variable, as well as the interaction effects for each pair of variables. They apply it to the prediction of both the 90 cross-rates quoted in both directions for the G10 currencies and a narrow set of five paired characteristics such as short-term interest rate differentials. They find that the

predicted linear effects of random forests, boosted trees, and NNs are very similar to a multiple linear regression. Subsequently, Li, Simon and Turkington (2022) apply the same methodology to the stocks of the S&P 500 using 16 predictors, which include both characteristics and a few macroeconomic variables. They implement several models to forecast returns: OLS, lasso, random forests, boosted trees and NNs, which they then combine with a trading rule that is long on the highest return forecasts, and short in the lowest ones. They find that nonlinearities are most important for boosted trees, while interactions play an important role in NNs. Daul, Jaisson and Nagy (2022) follow a similar methodology with 205 characteristics for international stocks, finding that the superiority of trees and NNs comes form their regularization and interactions rather than from the nonlinear component of individual predictions.

In turn, Jaeger et al. (2021) use Shapley values to assess the importance of the different predictor variables. Although they are nowadays prevalent in the machine learning literature, Shapley values were originally introduced in cooperative game theory to define a fair split of payouts across players considering their contributions to the common goal. In a prediction setting, one can interpret the predictors as the players and the prediction as the goal (see chapter 8 of Molnar (2022) and the references therein for additional details). Jaeger et al. (2021) study several strategies such as risk parity that focus on diversifying risk with 17 rolled-over futures on commodities, equity indexes, and fixed income. Specifically, they apply Shapley values to study the contribution of 96 statistics such as means, standard deviations or drawdowns in a boosted tree algorithm with the increment in Calmar ratios as their metric. The Calmar or drawdown ratio is a portfolio performance measure regularly applied to commodity trading advisors and hedge funds that compares the average annual rate of return for the last three years to the maximum peak to trough price fall over the same period. They conclude that the reduction in length and depth of drawdowns are the main drivers of the improvements that they observe.

Goulet Coulombe et al. (2023) develop an alternative methodology based on Shapley values for portfolio performance attribution across predictors. They apply their methodology to study which of the firm characteristics in Chen and Zimmermann (2022) are truly relevant for explaining the cross-section of stock returns. They consolidate 207 characteristics into 20 groups, combining a boosted tree for forecasts with long-short strategies based on quintile spreads. They conclude that pricing factors related to earnings, investment, and the seasonal "momentum" at annual frequencies studied by Heston and Sadka (2008) make relevant contributions to the Sharpe and Calman ratios of the portfolios.

The conventional wisdom in both econometrics and machine learning is that the optimal complexity of a model should appropriately balance the biases generated by simple models with

the variance associated to overparametrized ones. However, this widespread belief has been challenged recently. Specifically, in their study of double descent in deep learning, Schaeffer et al. (2023) argue that the usual bias-variance trade-off is limited to an initial region, with the prediction error going down again for higher levels of model complexity.

Similarly, Kelly, Malamud and Zhou (2023) use random matrix theory to highlight what they call the "virtue of complexity" in return prediction with ridge regularization. From the empirical point of view, they provide evidence of significant information ratios obtained by market timing strategies that use nonlinear transformations of the Goyal and Welch (2008) macroeconomic predictors. Interestingly, the successful market timing procedures that they derive learn to behave as long-only strategies that divest when they get close to macroeconomic recessions. Didisheim et al. (2023) extend the same analysis to SDF estimation for a large cross-section of individual stocks. In particular, they empirically study managed portfolios from nonlinear transformations of 130 characteristics, finding that the most successful factor pricing models are very large, using as many as one million factors.

# 4 Other impacts of big data in portfolio management

In this section we study two areas that are also important for portfolio management, but which have attracted less attention in the academic literature so far: the implications of big data for intertemporal portfolio decisions, and the use of alternative data, especially text and images.

## 4.1 Intertemporal portfolio decisions

To study optimal investment decisions in an intertemporal setting, we first review dynamic programming, the usual tool for making optimal sequential decisions in economics and finance. Afterwards, we discuss reinforcement learning, which is the branch of machine learning developed for those purposes, as it can accommodate big data more naturally.

### 4.1.1 Dynamic programming

As mentioned before, although conditional MV analysis is a natural extension of Markowitz' (1952) approach to a situation in which the mean or variance of asset returns are predictable, it is a myopic procedure in intertemporal contexts because it fails to take into account that next period investors will face the same problem all over again. In this respect, a basic lesson from dynamic programming is that a sequence of one-period solutions is not necessarily optimal when the investment horizon covers multiple periods.

One possibility would be to repeat the conditional MV analysis that we discus in section 3.1 for multiple horizons, as in Campbell and Viceira (2005), but again, this does not generally lead to internally consistent dynamic investment rules when portfolio rebalancing is allowed during the investment period. Some authors have proposed the growth-optimal portfolio, which is the one that maximizes the expected geometric return. However, this choice does not necessarily maximize the expected utility of final wealth for preferences different from the log function already considered by Bernoulli as a solution to the St. Petersburg paradox in the XVIII century.

Some practitioners use unconditional MV analysis to define what they call a strategic asset allocation (SAA), and conditional MV analysis to define what they call a tactical asset allocation (TAA). Tactical asset allocation, though, explicitly takes into account that the investor is holding the SAA portfolio as a background risk (see section 6 of Peñaranda (2008) for details).

Merton (1971) developed the first comprehensive solution to the intertemporal portfolio choice problem in a continuous time framework. He assumed not only that there is an instantaneously safe bond, but also that the log prices of a vector of risky assets that pay no dividends evolve according to a system of first-order stochastic differential equations driven by a multivariate Brownian motion process whose drifts and diffusions components depend on a set of $d$ predictor variables $\mathbf{x}$, which in turn evolve according to a first-order Markovian process. Specifically, if $d\mathbf{Y}$ denotes the vector whose entries are equal to $dP_i/P_i$ for each $i$, Merton (1971) assumed that

$$\begin{aligned} d\mathbf{Y} &= \boldsymbol{\nu}_{\mathbf{Y}}(\mathbf{x},t)dt + \boldsymbol{\Lambda}_{\mathbf{Y}}(\mathbf{x},t)d\mathbf{B}_{\mathbf{Y}}, \\ d\mathbf{x} &= \boldsymbol{\nu}_{\mathbf{X}}(\mathbf{x},t)dt + \boldsymbol{\Lambda}_{\mathbf{X}}(\mathbf{x},t)d\mathbf{B}_{\mathbf{X}}. \end{aligned}$$

Notice that by maintaining the assumption that the instantaneous conditional distribution of continuously compounded returns and predictor innovations is Gaussian and allowing for the existence of a conditionally riskless asset, Merton (1971) was effectively working with the continuous time analogue to Tobin's (1958) version of Markowitz (1952) MV model.

In this context, the risk premia, defined as the instantaneous mean of the vector of returns in excess of the instantaneous conditionally safe rate $R_0(\mathbf{x},t)$, will be given by

$$\boldsymbol{\mu} = E_t(d\mathbf{Y} - R\boldsymbol{\iota}_n dt) = (\boldsymbol{\nu}_{\mathbf{Y}} - R_0\boldsymbol{\iota}_n)dt,$$

while the instantaneous covariance matrix will be

$$\boldsymbol{\Sigma}dt = V_t(d\mathbf{Y} - R_0\boldsymbol{\iota}_n dt) = \boldsymbol{\Lambda}_{\mathbf{Y}}\boldsymbol{\Lambda}_{\mathbf{Y}}' dt.$$

The final ingredient is the instantaneous covariance between returns and predictor variables, which is given by

$$\mathbf{\Phi}dt{=}cov_t(d\mathbf{Y}{-}R_0\boldsymbol{\iota}_n dt, d\mathbf{x}) = \mathbf{\Lambda}_{\mathbf{Y}}(\mathbf{x},t)cov(d\mathbf{B}_{\mathbf{Y}}, d\mathbf{B}_{\mathbf{X}})\mathbf{\Lambda}'_{\mathbf{X}}(\mathbf{x},t).$$

Let $W(t)$ denote the investor wealth at time $t$. The intertemporal portfolio problem consists in choosing the dynamic portfolio strategy $\mathbf{w}(W,\mathbf{x},t)$ that solves the following dynamic programme

$$\max_{\mathbf{w}} E_0\{u[W(T)]\} \quad s.t. \quad dW = W[(R_0 + \mathbf{w}'\boldsymbol{\mu})dt + \mathbf{w}'\mathbf{\Lambda}_{\mathbf{Y}}(\mathbf{x},t)d\mathbf{B}_{\mathbf{Y}}]$$

for some initial wealth $W(0) > 0$. The problem can be generalized to include utility over consumption between 0 and $T$ rather than over final wealth, as well as an infinite investment horizon in which $T \to \infty$.

Let us define the value function as the optimal expected utility at an intermediate period $t$, namely

$$V(W,\mathbf{x},t) = \max_{\mathbf{w}} E_t\{u[W(T)]\}.$$

The optimal value of the dynamic portfolio weights $\mathbf{w}$ is given by

$$\lambda_m \mathbf{\Sigma}^{-1}\boldsymbol{\mu} + \mathbf{\Sigma}^{-1}\mathbf{\Phi}\boldsymbol{\lambda}_h, \tag{18}$$

where

$$\lambda_m = -\frac{\partial V(W,\mathbf{x},t)/\partial W}{W\partial^2 V(W,\mathbf{x},t)/(\partial W)^2}$$

and

$$\boldsymbol{\lambda}_h = -\frac{\partial^2 V(W,\mathbf{x},t)/\partial \mathbf{x}\partial W}{W\partial^2 V(W,\mathbf{x},t)/(\partial W)^2}.$$

In general, the optimal portfolio weights in (18) depend on $W$, $\mathbf{x}$ and $t$, albeit $W$ becomes irrelevant when the utility function of the investor belongs to the CRRA class.

In this context, the optimal portfolio rule is the sum of two components:

1. the myopic conditional MV portfolio whose weights are $\mathbf{\Sigma}^{-1}\boldsymbol{\mu}$.

2. the so-called hedging demands, which protect the investor from the effects of unexpected changes in the predictor variables $\mathbf{x}$ by selecting the weights $\mathbf{\Sigma}^{-1}\mathbf{\Phi}$ of $d$ portfolios that most closely approximate their innovations in the usual mean square error sense.

The dynamic portfolio rule in (18) reduces to the myopic one in some important cases:

1. when returns are *i.i.d.* so that $R_0$, $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ do not depend on $\mathbf{x}$.

2. when utility is logarithmic because then $\partial^2 V(W, \mathbf{x}, t)/\partial \mathbf{x} \partial W = \mathbf{0}$, in which case the solution coincides with the growth-optimal portfolio.

3. when $\mathbf{\Phi} = \mathbf{0}$, so that the risky assets cannot be used to hedge the shocks to the predictor variables.

Despite its elegance, Merton's (1971) approach is difficult to implement. Brennan, Schwartz and Langado (1997) provide a tractable solution when the drifts of the diffusions are a function of a small number of state variables. In turn, Campbell and Viceira (1999, 2001) obtain the optimal portfolio in a discrete time model in which the asset returns and the predictor variables obey a conditionally homoskedastic Gaussian vector autoregressive process of order 1 or VAR(1), while Chacko and Viceira (2005) do the same in a model with stochastic volatility.

Inspired by Johannes, Korteweg and Polson (2014), Babiak and Baruník (2021) apply deep learning forecasts of stock returns to long-horizon dynamic portfolio choice. Their objective is the market timing of the S&P 500 using 12 macro predictors from Goyal and Welch (2008). For a power utility investor, they find statistically and economically significant increases in certainty equivalent returns and Sharpe ratios when using deep learning methods. In addition, they find that an LSTM network of the type we discuss in section 4.2 outperforms feedforward networks. The benefits that they find also extend to other important performance characteristics such as drawdowns even after considering transaction costs.

Garleanu and Pedersen (2013) suggested an alternative intertemporal decision framework in which agents care about the present discounted value of intraperiod MV preferences in the presence of quadratic transaction costs when expected returns are linear combinations of some state variables that follow another conditionally homoskedastic Gaussian VAR(1). Under certain restrictions, the dynamic programming solution is a convex combination of the current optimal MV portfolio and what they call the target MV portfolio, whose weights depend on the sensitivity of each asset to the state variables and the persistence of these variables.

More recently, Jensen et al. (2022) have proposed a variation of this framework in a stationary environment that allows for general forms of dependence of the current and future expected returns on the state variables. Although they find that a tractable approximation to the optimal strategy involves both the previous optimal weights and a combination of the current and expected future MV portfolios, they recommend to learn directly the optimal strategy using random Fourier features, which is a machine learning technique related to sieves in which the basis functions are randomly chosen. They work with 115 characteristics for individual US stocks, and consider a monthly portfolio rebalancing. Their methods produce a better out-of-sample MV frontier net of trading costs. Jensen et al. (2022) also contribute to the interpretability

of machine learning methods by providing an economic measure that captures the contribution of each predictor to realized utility. Specifically, they find that transaction costs increase the economic relevance of persistent features related to value, quality, and momentum.

### 4.1.2 Reinforcement learning

In previous sections, we have studied various portfolio management applications of what data scientists call supervised and unsupervised learning. In this section, in contrast, we study applications of a third branch of machine learning called reinforcement learning, a data-driven approach to sequential decision making. This family of methods provides a solution in a context similar to the one in section 4.1.1, but the main difference is that the decision maker does not necessarily know either the transition probability between current and future states or even her own utility function. Under certain conditions, a stochastic learning policy that considers both "exploration" by means of random action and "exploitation" of the action that currently maximizes rewards converges to the optimal policy that she would apply in a standard dynamic programming framework with perfect knowledge of her utility function and the transition probabilities of the Markov processes involved.[10]

In this respect, Chaouki et al. (2020) study several dynamic portfolio choice set-ups and find that overall reinforcement learning can recover the essential features of the corresponding optimal strategies, thereby achieving a close-to-optimal performance. More generally, Garcia and Marinenko (2024) review reinforcement learning for portfolio allocation. In turn, chapter 10, section 5 of Dixon, Halperin and Bilokon (2020) considers the application of reinforcement learning to stock portfolios, while section 6 studies wealth management, providing an explicit connection to Merton (1971).

Moody et al. (1998) is an early reference of market timing with the S&P 500. They use several portfolio evaluation measures such as terminal wealth, utility or Sharpe ratio, finding that a long-short trader can obtain better results with reinforcement learning than with a supervised learning procedure focused on forecasting returns.

Wang and Zhou (2020) work in a continuous-time framework where returns are *i.i.d.* and the investor has MV preferences over terminal wealth, but does not know the model parameters. In their empirical study with a ten-year investment horizon and monthly rebalancing for sets of twenty stocks, they find that their reinforcement learning algorithms, which do not use NNs, yield

[10] Two of the best known examples of reinforcement learning are as follows. In 2013, DeepMind showed that it was able to learn to play most Atari video games using pixel data only without any previous knowledge of the rules. A few years later, these methods managed to beat human masters in the game Go. More recently, reinforcement learning methods that use deep NNs in their implementation when the number of states and actions is large have become crucial ingredients in other applications more relevant in real life, such as self-driving cars, and recommendation systems for books, films and songs.

better results than an MV portfolio, an equally weighted one, and an alternative reinforcement learning algorithm that requires deep NNs.

Recently, Cong et al. (2022) use deep reinforcement learning for monthly rebalancing of US stock portfolios when the conditioning information consists of 51 characteristics along the lines of Freyberger, Neuhierl and Weber (2020). They use a network architecture similar to the ones we describe in section 4.2, which has three components: a transformer encoder for each asset's time-series data, a cross-section network to describe relationships across assets, and the construction of a portfolio with long and short positions for assets with high and low winner scores, respectively, obtained from the first and second components.[11] They train their model using a reinforcement learning procedure in which the reward is a performance measure such as the Sharpe ratio. They report high Sharpe ratios and alphas, and also provide some interpretability with gradient-based methods and lasso, finding that rotation and nonlinearity are the key drivers of portfolio performance.

### 4.2 Alternative data

In this last section, we review several portfolio management applications that successfully exploit alternative data sources, an area in which professional investors such as Renaissance Technologies have been substantially ahead of academic researchers for a long time. We focus on textual and still image data because they have been more frequently used so far, but we expect audio and video to play an increasingly important part of the investment process in the future, as improvements in machine learning techniques simplify their use.

The analysis of textual data obtained from news sources, central bank monetary policy statements, official company filings, and social media has been by far the most widely used source of alternative data in empirical work (see chapters 4 to 6 in Cao (2023) for applications of textual data to financial investing with a practitioner perspective).

Tetlock (2007) and Fang and Peress (2009) convincingly highlighted the predictive power of media coverage for stock returns. Zhang and Skiena (2010) is another early application of sentiment analysis in the development of trading strategies. They apply a predecessor of the natural language processing techniques described below to several blogs and news sources to identify references to 3,238 individual US companies traded on the NYSE and their relationships. On this basis, they create time series of favorable or unfavorable words for each of them, which they summarize by means of a "polarity" indicator defined as the fraction of positive sentiment references among total sentiment references. Their market-neutral trading strategy,

[11] See Cong et al. (2021) for additional empirical evidence of transformers, LSTMs, and other NNs with the commonly used decile spreads.

which goes long in stocks with positive sentiment and short in those with negative one, is able to systematically improve upon the performance of both its mirror-image strategy and a third selection strategy that chooses long and short positions at random.

More recently, Agrawal et al. (2018) also study the relationship between news and social-media sentiment indicators they obtain from StockTwits and Twitter messages and 500 individual stock returns, trading volume and liquidity. Based on their empirical results, they consider the following intraday mean-reversion trading strategy: every 30 minutes, they buy equities that had negative returns over the previous interval and short-sell those that had positive ones, giving higher absolute weights to those companies with a larger number of social-media messages. They show that using these tilts to the portfolio weights based on social media outperforms an analogous mean-reversion strategy that only uses price reversions.

At the level of the aggregate stock market, Azar and Lo (2016) construct a daily polarity sentiment indicator based on tweets issued in anticipation of Federal Reserve meetings to predict the reaction of the S&P 500 to announcements by the Federal Open Market Committee, showing that an MV market timing similar to the one discussed in section 3.1 in which expected excess returns depend on the polarity score outperforms a passive buy-and-hold strategy. Similarly, Karagozoglu and Fabozzi (2017) make use of a related "wisdom of crowds" approach with a broader social media coverage to predict shifts in stock market volatility, which they successfully exploit by trading Exchange Traded Futures and Notes related to the VIX, the most popular volatility index.

In turn, Cohen, Malloy and Nguyen (2020) obtain 10-Q and 10-K filings of US firms, two reports about the financial performance of publicly traded companies submitted by these to the US Securities and Exchange Commission on a quarterly and annual basis, respectively, focusing on the content of these documents along the lines of Loughran and McDonald (2011). They work with sentiment identifiers from the master dictionary in that paper, as well as with analyst data. They then compute several measures of similarity between 10-Q and 10-K filings for successive periods, such as cosine similarity across terms, a correlation-type measure based on the inner product of their vector representations, and for each of them they construct monthly portfolios based on quintile spreads. These authors report that a portfolio that takes long positions in firm that do not substantially modify their annual and quarterly reports and short positions in those that do earns a positive alpha.

Recently, Bybee, Kelly and Su (2023) develop a method to estimate the state variables and asset pricing factors underlying Merton's (1973) Intertemporal CAPM (ICAPM) from business news narratives. In particular, they work with several decades of the full text of the Wall Street

Journal (WSJ) and the corresponding daily narrative attention estimates based on topic modeling to decrease the dimensionality of the article contents to 180 topics.[12] They use a variant of the latent factor analysis of Kelly, Pruitt and Su (2019) that generalizes Fama-MacBeth regressions to deal with unobserved state variables related to those topics using a group lasso penalty for the selection of the relevant narratives. Their estimated "narrative" pricing factors achieve a higher Sharpe ratio than the Fama-French and momentum factors together, and successfully predict future investment opportunities along the lines of the ICAPM. In addition, they manage to find plausible interpretations of the narrative factors, with the "recession" narrative having the largest impact on the estimated SDF.

All the previous papers rely on the traditional "bag-of-words" approach, which only looks at the frequency of words rather than their ordering in a sequence. More modern approaches in natural language processing work with vector representations of the text called "embeddings". In addition, during the last few years there has been an important change in the default architecture, which has moved away from recurrent neural networks (RNN) such as LSTM towards transformers such as BERT and GPT commonly known as large language models (LLM). In this respect, ChatGTP, a famous chatbot and virtual assistant based on the GTP transformer, has recently captured people's imagination. RNNs process the sequences of textual data recursively because at each time step both the input and the outcome from the previous time step are considered, analogously to a nonlinear autoregressive model with an exogenous predictor. In contrast, transformers can process data in parallel, which lets them handle much larger datasets while keeping information on the position of each element of a sequence and providing a contextualized representation of textual data.

Chen, Kelly and Xiu (2023) work with embeddings from both word-based models and LLMs such as BERT. These authors input their textual data into open-source, pre-trained models to generate embeddings, which they then use as inputs to their sentiment analysis and return forecasting. They obtain global news in thirteen different languages from Refinitiv, a global provider of financial data, and work with stocks from 16 international markets. Lastly, they construct quintile spreads from sentiment scores with a daily rebalance, showing that their portfolio strategies achieve high Sharpe ratios.

Lopez-Lira and Tang (2023) create a sentiment score with ChatGPT applied to news headlines from various sources, which they match with those of the commercial sentiment analysis provider RavenPack. They build a daily strategy that goes long in stocks with positive score,

[12] These estimates are taken from Bybee et al. (2023), who study the WSJ corpus in its "bag-of-words" form. They also perform a market timing exercise with positions proportional to the return forecasts from a lasso regression.

and short in stocks with negative one, which they complement with some interpretability analysis. In turn, Chen et al. (2023) apply ChatGPT in a monthly market-timing exercise. They extract the proportions of good and bad stock market news from the WSJ, and use them to forecast returns, which they then use in a conditional MV portfolio. Both papers find considerable improvements by using ChatGPT instead of previous LLMs such as BERT.

Before the current textual data revolution took off, though, the main driver behind the adoption of deep learning methods was the analysis of image data in computer vision. In particular, convolutional neural networks (CNN) proved very powerful in extracting information from images. These are specialized networks for processing data with a known grid structure. Examples include equidistant time-series data (a 1D grid) and image data (a 2D or 3D grid of pixels depending on the use of a color scale).

In the specific context of portfolio management, chapter 13 in Denev and Amen (2020) describes for practitioners several applications of image data obtained from security cameras, drones and satellites, including the measurement of nighttime light intensity, changes in land cover classification, as well as retail parking lots use.

In an exercise that combines text and images, Obaida and Pukthuanthong (2022) collect 148,823 articles and their associated photos from the WSJ, re-estimating (or "fine-tuning") the last layer of the pre-trained Google Inception model with the latter in an example of transfer learning. They employ the DeepSent data for that purpose, which has photos labeled by sentiment. Using these methods, they build a daily sentiment index with the proportion of photos predicted to have negative sentiment each day. They also build a similar indicator for the article headline and summary, finding that the images convey alternative information to the text. In addition, they compute the certainty equivalent return and the Sharpe ratio of a daily, conditional MV, market timing strategy in which the return forecast is linear in the photo pessimism, and the variance forecast is obtained over an expanding window of past returns. They find notably higher performance than by simply relying on return forecasts based on historical averages.

Finally, drawing inspiration from technical analysis, Jiang, Kelly and Xiu (2023) extract buy and sell signals from price-charts for individual stocks, but using a CNN prediction model. Specifically, they rank individual stocks according to the probability that their return is positive in the next period, on the basis of which they construct the usual long-short decile spread portfolio. They find that the return forecast from a CNN is distinctly different from traditional price trend signals, outperforming trading strategies based on them. They also find that the CNN trained on US daily data offers transfer learning opportunities for both lower frequency data and foreign equities.

# References


Agrawal, S., P.D. Azar, A.W. Lo and T. Singh (2018): "Momentum, mean-reversion, and social media: evidence from StockTwits and Twitter", *Journal of Portfolio Management* 44, 85–95.

Aït-Sahalia, Y. and J. Jacod (2014): *High-frequency financial econometrics*, Princeton University Press.

Aït-Sahalia, Y. and D. Xiu (2017): "Using principal component analysis to estimate a high dimensional factor model with high-frequency data", *Journal of Econometrics* 201, 384-389.

Andersen, T. G., T. Bollerslev and F. X. Diebold (2010): "Parametric and nonparametric measurement of volatility", in Y. Aït-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics* 67–138, North Holland.

Ao, M., Y. Li and X. Zheng (2019): "Approaching mean-variance efficiency for large portfolios", *Review of Financial Studies* 32, 2890–2919.

Asness, C.S., T.J. Moskowitz and L.H Pedersen (2013): "Value and momentum everywhere", *Journal of Finance*, 68, 929–985.

Avramov, D., S. Cheng and L. Metzker (2023): "Machine learning versus economic restrictions: evidence from stock return predictability", *Management Science* 69, 2547–3155.

Avramov, D. and G. Zhou (2010): "Bayesian portfolio analysis", *Annual Review of Financial Economics* 2, 25–47.

Azar, P.D. and A.W. Lo (2016): "Practical applications of the wisdom of Twitter crowds: predicting stock market reactions to FOMC meetings via Twitter", *Journal of Portfolio Management* 42, 123–134.

Babiak, M. and J. Baruník (2021): "Deep learning, predictability, and optimal portfolio returns", https://arxiv.org/pdf/2009.03394.

Bai, J. (2003): "Inferential theory for factor models of large dimensions", *Econometrica* 71, 135-171.

Bai, J. and S. Ng (2002): "Determining the number of factors in approximate factor models", *Econometrica* 70, 191–221.

Banz, R.W. (1981): "The relationship between return and market value of common stocks", *Journal of Financial Economics* 9, 3–18.

Barndorff-Nielsen, O. E. and N. Shephard (2007): "Variation, jumps and high frequency data in financial econometrics", in R. Blundell, T. Persson and W. K. Newey (eds.), *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*, Econometric Society Monographs, 328–372. Cambridge University Press.

Barras, L., O. Scaillet and R. Wermers (2010): "False discoveries in mutual fund performance:

measuring luck in estimated alphas", *Journal of Finance* 65, 179–216.

Barras, L., O. Scaillet and R. Wermers (2018): "False discoveries in mutual fund performance: measuring luck in estimated alphas: a reply", *Journal of Finance* Replication and Corrigenda https://afajof.org/wp-content/uploads/22200324-Reply-for-JF.pdf .

Bartram, S.M., J. Branke and M. Motahari (2020): "Artificial intelligence in asset management", CFA Institute Research Foundation.

Bawa, V. S., S. J. Brown and R.W. Klein (1979): *Estimation risk and optimal portfolio choice*, North Holland.

Best, M.J. and R.R. Grauer (1991): "On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results", *Review of Financial Studies* 4, 315-342.

Bianchi, D., M. Büchner and A. Tamoni (2021): "Bond risk premiums with machine learning", *Review of Financial Studies* 32, 1046–1089.

Black, F. and R. Litterman (1992): "Global portfolio optimization", *Financial Analysts Journal* 48, 28-43.

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

Brandt, M.W., P. Santa-Clara and R. Valkanov (2009): "Parametric portfolio policies: exploiting characteristics in the cross-section of equity returns", *Review of Financial Studies* 22, 3411–3447.

Brennan, M.J., E.S. Schwartz and R. Lagnado (1997): "Strategic asset allocation", *Journal of Economic Dynamics and Control* 21, 1377-1403.

Bryzgalova, S., V. De Miguel, S. Li and M. Pelger (2023): "Asset-pricing factors with economic targets", https://ssrn.com/abstract=4344837.

Bryzgalova, S., M. Pelger and J. Zhu (2023): "Forest through the trees: building cross-sections of stock returns", forthcoming in the *Journal of Finance*.

Bybee, L., B. Kelly, A. Manela and D. Xu (2023): "Business news and business cycles", forthcoming in the *Journal of Finance*.

Bybee, L., B. Kelly and Y. Su (2023): "Narrative asset pricing: interpretable systematic risk factors from news text", *Review of Financial Studies* 36, 4759–4787.

Campbell, J.Y. and L.M. Viceira (1999): "Consumption and portfolio decisions when expected returns are time varying", *Quarterly Journal of Economics* 114, 433-495.

Campbell, J.Y. and L.M. Viceira (2001): "Who should buy long-term bonds?", *American Economic Review* 91, 99-127.

Campbell, J.Y. and L.M. Viceira (2005): "The term structure of the risk-return trade-off",

*Financial Analysts Journal* 61, 34-44.

Cao, K. and H. You (2024): "Fundamental analysis via machine learning", forthcoming in the *Financial Analyst Journal*.

Cao, L. (ed.) (2023): *Handbook of artificial intelligence and big data applications in investments*. CFA Institute Research Foundation.

Capponi, A. and C-A. Lehalle (eds.) (2023): *Machine learning and data sciences for financial markets*. Cambridge University Press.

Carhart, M.M. (1997): "On persistence in mutual fund performance", *Journal of Finance* 52, 57-82.

Chamberlain, G. (1983): "Funds, factors and diversification in arbitrage pricing models", *Econometrica* 51, 1305-1324.

Chamberlain, G. and M. Rothschild (1983): "Arbitrage, factor structure, and mean-variance analysis on large asset markets", *Econometrica* 51, 1281-1304.

Chacko, G. and L.M. Viceira (2005): "Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets", *Review of Financial Studies* 18, 1369-1402.

Chaouki, A., S. Hardiman, C. Schmidt, E. Serie and J. de Lataillade (2020): "Deep deterministic portfolio optimization", *Journal of Finance and Data Science* 6, 16-30.

Chen, A.Y. and M. Velikov (2023): "Zeroing in on the expected returns of anomalies", *Journal of Financial and Quantiative Analysis* 58, 968–1004.

Chen, A.Y. and T. Zimmermann (2022): "Open source cross-sectional asset pricing", *Critical Finance Review* 11, 207–264.

Chen, J, G. Tang, G. Zhou and W. Zhu (2023): "ChatGPT, stock market predictability and links to the macroeconomy", Olin Business School Center for Finance & Accounting Research Paper 2023/18.

Chen, L., M. Pelger and J. Zhu (2024): "Deep learning in asset pricing", *Management Science* 70, 671–1342.

Chen, X. (2007): "Large sample sieve estimation of semi-nonparametric models", in J. Heckman and E. Leamer (eds.), *Handbook of Econometrics* vol. 6B, chapter 76, Elsevier.

Chen, X., F. Peñaranda, D. Pouzo and E. Sentana (2023): "Sieve managed portfolios", mimeo, CEMFI.

Chen, Y., B. Kelly and D. Xiu (2023): "Expected returns and large language models", https://ssrn.com/abstract=4416687.

Choi, D. W. Jiang and C. Zhang (2024): "Alpha go everywhere: machine learning and international stock returns", https://ssrn.com/abstract=3489679.

Cohen, L., C. Malloy and Q. Nguyen (2020): "Lazy prices", *Journal of Finance* 75, 1371–1415.

Cong, L.W., K. Tang, J. Wang and Y. Zhang (2021): "Deep sequence modeling: development and applications in asset pricing", *Journal of Financial Data Science* 3, 28–42.

Cong, L.W., K. Tang, J. Wang and Y. Zhang (2022): "AlphaPortfolio: direct construction through deep reinforcement learning and interpretable AI", https://ssrn.com/abstract=3554486.

Cong, L.W., G. Feng, J. He and X. He (2023): "Growing the efficient frontier on panel trees", NBER Working Paper 30805.

Connor, G. and R. Korajczyk (1988): "Risk and return in an equilibrium APT: application of a new testing methodology", *Journal of Financial Economics* 21, 255-289.

Connor, G. and R. Korajczyk (1993): "A test for the number of factors in an approximate factor model", *Journal of Finance* 48, 1263-1291.

Da, R., S. Nagel and D. Xiu (2023): "The statistical limit of arbitrage", mimeo, University of Chicago.

Daul, S., T. Jaisson and A. Nagy (2022): "Performance attribution of machine learning methods for stock return predictions", *Journal of Finance and Data Science* 8, 86–104.

Dello Preite, M., R. Uppal, P. Zaffaroni and I. Zviadadze (2024): "Cross-sectional asset pricing with unsystematic risk", https://ssrn.com/abstract=4135146.

De Bondt, W.F.M. and R. Thaler (1984): "Does the stock market overreact?", *Journal of Finance* 40, 793-805.

De Miguel, V., L. Garlappi and R. Uppal (2009): "How inefficient is the 1/N asset-allocation strategy?", *Review of Financial Studies* 22, 1915–1953.

De Miguel, V., A. Martín-Utrera, F.J. Nogales and R. Uppal (2020): "A transaction-cost perspective on the multitude of firm characteristics", *Review of Financial Studies* 33, 2180–2222.

Denev, A. and S. Amen (2020): *The book of alternative data: a guide for investors, traders, and risk managers.* Wiley.

Didisheim, A., S. Ke, B. Kelly and S. Malamud (2023): "Complexity in factor pricing models", NBER Working Paper 31689.

Dixon, M.F., I. Halperin and P. Bilokon (2020): *Machine learning in finance: from theory to practice.* Springer.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004): "Least angle regression", *Annals of Statistics* 32, 407-499.

Elton, E.J. and M.J. Gruber (1973): "Estimating the dependence structure of share prices", *Journal of Finance* 28, 1203–1232.

Fabozzi, F. J., D. Huang and G. Zhou (2010): "Robust portfolios: contributions from operations

research and finance", *Annals of Operations Research* 176, 191-220.

Fama, E.F. and K.R. French (1993): "Common risk factors in the returns on stock and bonds", *Journal of Financial Economics* 33, 3-56.

Fama, E.F. and K.R. French (2015): "A five-factor asset pricing model", *Journal of Financial Economics* 116, 1–22.

Fama, E.F. and J.D. MacBeth (1973): "Risk, return, and equilibrium: empirical tests", *Journal of Political Economy* 81, 607–636.

Fan, J., Y. Fan and J. Lv (2008): "High dimensional covariance matrix estimation using a factor model", *Journal of Econometrics* 147, 186–197.

Fang, L. and Peress, J. (2009): "Media coverage and the cross-section of stock returns", *Journal of Finance* 64, 2023–2052.

Freyberger, J., A. Neuhierl and M. Weber (2020): "Dissecting characteristics nonparametrically", *Review of Financial Studies* 33, 2326–2377.

Friedman, J.H. (2001): "Greedy function approximation: a gradient boosting machine", *Annals of Statistics* 29, 1189-1232.

Frost, P.A. and Savarino, J.E. (1986): "An empirical Bayes approach to efficient portfolio selection", *Journal of Financial and Quantitative Analysis* 21, 293-305.

Gallant, A.R., L.P. Hansen and G. Tauchen (1990): "Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution", *Journal of Econometrics* 45, 141-179.

Garcia, R. and A. Marinenko (2024): "Portfolio allocation and reinforcement learning", in M. Corazza, R. Garcia, F. Khan, D. LaTorre and H. Masri (eds.), *Artificial intelligence and beyond in finance*, World Scientific.

Garleanu, N. and L.H. Pedersen (2013): "Dynamic trading with predictable returns and transaction costs", *Journal of Finance* 63, 2309-2340.

Giglio, S., Y. Liao and D. Xiu (2021): "Thousands of alpha tests", *Review of Financial Studies* 34, 3456–3496.

Giglio, S., B. Kelly and D. Xiu (2022): "Factor models, machine learning, and asset pricing", *Annual Review of Financial Economics* 14, 337–368.

Goetzmann, W., J. Ingersoll, M. Spiegel and I. Welch (2007): "Portfolio performance manipulation and manipulation-proof performance measures", *Review of Financial Studies* 20, 1503–1546.

Goulet Coulombe, P. and M. Göbel (2023): "Maximally machine-learnable portfolios", https://arxiv.org/abs/2306.05568.

Goulet Coulombe, P., D.R. Rapach, E.C. Montes Schütte and S. Schwenk-Nebbe (2023): "The

anatomy of machine learning-based portfolio performance", https://ssrn.com/abstract=4628462.

Goyal, A. and I. Welch (2008): "A comprehensive look at the empirical performance of equity premium prediction", *Review of Financial Studies* 21, 1455–1508.

Goyal, A., I. Welch and A. Zafirov (2023): "A comprehensive 2022 look at the empirical performance of equity premium prediction", https://ssrn.com/abstract=3929119.

Gu, S., B. Kelly and D. Xiu (2020): "Empirical asset pricing via machine learning", *Review of Financial Studies* 33, 2223–2273.

Guida, T. (Ed.) (2019): *Big data and machine learning in quantitative investment*. Wiley.

Guidolin, M. (2024): "Machine learning in portfolio decisions", in M. Corazza, R. Garcia, F. Khan, D. LaTorre and H. Masri (eds.) *Artificial Intelligence and Beyond in Finance*, World Scientific.

Han, Y., A. He, D.E. Rapach and G. Zhou (2021): "Expected stocks returns and firm characteristics: E-LASSO, assessment, and implications", mimeo, Saint Louis University.

Hansen, L.P. and R. Jagannathan (1991): "Implications of security market data for models of dynamic economies", *Journal of Political Economy* 99, 225-262.

Hansen, L.P. and S.F. Richard (1987): "The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models", *Econometrica* 55, 587–613.

Harvey, C.R. and Y. Liu (2020): "False (and missed) discoveries in financial economics", *Journal of Finance* 75, 2503–2553.

Harvey, C.R., Y. Liu and H. Zhu (2015): "... and the cross-section of expected returns", *Review of Financial Studies* 29, 5–68.

Heston, S.L. and R. Sadka (2008): "Seasonality in the cross-section of stock returns", *Journal of Financial Economics* 87, 418–445

Hull, J.C. (2021): *Machine learning in business: an introduction to the world of data science*, 3rd edition, independently published.

Israel, R., B. Kelly and T. Moskowitz (2020): "Can machines "learn" finance?", *Journal of Investment Management* 18, 23–36.

Jacobsen, B., F. Jiang and H. Zhang (2019): "Ensemble machine learning and stock return predictability", mimeo, Tilburg University.

Jaeger, M., S. Krügel, D. Marinelli, J. Papenbrock and P. Schwendner (2021): "Interpretable machine learning for diversified portfolio construction", *Journal of Financial Data Science* 3, 31–51.

Jagannathan, R. and T. Ma (2003): "Risk reduction in large portfolios: why imposing the wrong constraints helps", *Journal of Finance* 58, 1651-1684.

Jagannathan, R. and Z. Wang (1996): "The conditional CAPM and the cross-section of expected returns", *Journal of Finance* 51, 3-53.

Jegadeesh, N. and S. Titman (1993): "Returns to buying winners and selling losers: implications for stock market efficiency", *Journal of Finance* 48, 65-91.

Jensen, T.I., B. Kelly, S. Malamud and L.H. Pedersen (2022): "Machine learning and the implementable efficient frontier", Swiss Finance Institute Research Paper 22-63.

Jiang, J., B. Kelly and D. Xiu (2023): "(Re-)Imag(in)ing price trends", *Journal of Finance* 78, 3193–3249.

Johannes, M., A. Korteweg and N. Polson (2014): "Sequential learning, predictability, and optimal portfolio returns", *Journal of Finance* 69, 611–644.

Jorion, P. (1986): "Bayes-Stein estimation for portfolio analysis", *Journal of Financial and Quantitative Analysis* 21, 278–292.

Kan, R. and G. Zhou (2007): "Optimal portfolio choice with parameter uncertainty", *Journal of Financial and Quantitative Analysis* 42, 621–656.

Karagozoglu, A.K. and F.J. Fabozzi (2017): "Volatility wisdom of social media crowds", *Journal of Portfolio Management* 43, 136–151.

Kelly, B., S. Malamud and L.H. Pedersen (2023): "Principal portfolios", *Journal of Finance* 78, 347–387.

Kelly, B., S. Malamud and K. Zhou (2023): "The virtue of complexity in return prediction", *Journal of Finance* 79, 459–503.

Kelly, B., D. Palhares and S. Pruitt (2023): "Modeling corporate bond returns", *Journal of Finance* 78, 1967–2008.

Kelly, B., S. Pruitt and Y. Su (2019): "Characteristics are covariances: A unified model of risk and return", *Journal of Financial Economics* 134, 501–524.

Kelly, B. and D. Xiu (2023): "Financial machine learning", *Foundations and Trends in Finance* 13, 205–363.

Kim, S., R.A. Korajczyk and A. Neuhierl (2021): "Arbitrage portfolios", *Review of Financial Studies* 34, 2813–2856.

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

Ledoit, O. and M. Wolf (2003): "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection", *Journal of Empirical Finance* 10, 603-621.

Leipold, M., Q. Wang and W. Zhou (2022): "Machine learning in the Chinese stock market", *Journal of Financial Economics* 145, 64–82.

Lehmann, B.N. and D. Modest (1988): "The empirical foundations of the arbitrage pricing theory", *Journal of Financial Economics* 21, 213-254.

Lettau. M. and M. Pelger (2020a): "Estimating latent asset-pricing factors", *Journal of Econometrics* 218, 1–31.

Lettau. M. and M. Pelger (2020b): "Factors that fit the time series and cross-section of stock returns", *Review of Financial Studies* 33, 2274–2325.

Li, Y., Z. Simon and D. Turkington (2022): "Investable and interpretable machine learning for equities", *Journal of Financial Data Science* 4, 54–74.

Li, Y., D. Turkington and A. Yazdani (2020): "Beyond the black box: an intuitive approach to investment prediction with machine learning", *Journal of Financial Data Science* 2, 61–75.

Lo, A.W. and A.C. Mackinlay (1997): "Maximizing predictability in the stock and bond markets", *Macroeconomic Dynamics* 1, 102-134.

López de Prado, M. (2018): *Advances in financial machine learning.* Wiley.

López de Prado, M. (2020): *Machine learning for asset managers.* Cambridge University Press.

López-Lira, A. and Y. Tang (2023): "Can ChatGPT forecast stock price movements? Return predictability and large language models", https://arxiv.org/abs/2304.07619.

Loughran, T. and B. McDonald (2011): "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks", *Journal of Finance* 66, 35-65.

Maillard, S., T. Roncalli and Teïletche, J. (2010): "On the properties of equally-weighted risk contributions portfolios", *Journal of Portfolio Management* 36, 60-70.

Manresa, E., F. Peñaranda and E. Sentana (2023): "Empirical evaluation of overspecified asset pricing models", *Journal of Financial Economics* 147, 338-351.

Markowitz, H. (1952): "Portfolio selection", *Journal of Finance* 8, 77-91.

McCracken, M. W. and S. Ng (2016): "FRED-MD: A monthly database for macroeconomic research", *Journal of Business and Economic Statistics* 34, 574–589.

Merton, R.C. (1971): "Optimum consumption and portfolio rules in a continuous-time model", *Journal of Economic Theory* 3, 373-413.

Merton, R. C. (1973): "An intertemporal capital asset pricing model", *Econometrica* 41, 867–887.

Merton, R.C. (1980): "On estimating the expected return on the market: an exploratory investigation", *Journal of Financial Economics* 8, 323-361.

Mirete-Ferrer, P.M., A. García-García, J.S. Baixauli-Soler and M.A. Prats (2022): "A review on machine learning for asset management", *Risks* 10, 84.

Molnar (2022): *Interpretable machine learning*, independently published.

Moody, J., L. Wu, Y. Liao and M. Saffell (1998): "Performance functions and reinforcement learning for trading systems and portfolios", *Journal of Forecasting* 17, 441-470.

Moritz, B. and T. Zimmermann (2016): "Tree-based conditional portfolio sorts: the relation between past and future stock returns", https://ssrn.com/abstract=2740751.

Obaida, K. and K. Pukthuanthong (2022): "A picture is worth a thousand words: measuring investor sentiment by combining machine learning and photos from news", *Journal of Financial Economics* 144, 273–297.

Onatski, A. (2012): "Asymptotics of the principal components estimator of large factor models with weakly influential factors", *Journal of Econometrics* 168, 244-258.

Pástor L. (2000): "Portfolio selection and asset pricing models", *Journal of Finance* 55, 179–223

Pástor L and R.F. Stambaugh (2000): "Comparing asset pricing models: an investment perspective", *Journal of Financial Economics* 56, 335–81.

Peña, D and R.S Tsay (2021): *Statistical learning for big dependent data.* Wiley.

Peñaranda (2008): "Portfolio choice beyond the traditional approach", *Revista de Economía Financiera* 15, 50–90.

Peñaranda (2016): "Understanding portfolio efficiency with conditioning information", *Journal of Financial and Quantitative Analysis* 51, 985–1011.

Peñaranda, F. and E. Sentana (2011): "Inferences about return and stochastic discount factor mean-variance frontiers", mimeo, CEMFI.

Peñaranda, F. and E. Sentana (2015): "A unifying approach to the empirical evaluation of asset pricing models", *Review of Economics and Statistics* 97, 412-435.

Peñaranda, F. and E. Sentana (2016): "Duality in mean-variance frontiers with conditioning information", *Journal of Empirical Finance* 38, 762–785.

Peñaranda, F. and L. Wu (2022): "Targets, predictability, and performance", *Management Science* 68, 809–1589.

Rasekhschaffe, K.C. and R.C. Jones (2019): "Machine learning for stock selection", *Financial Analysts Journal* 75, 70–88.

Rosenberg, B., K. Reid and R. Lanstein (1985): "Persuasive evidence of market inefficiency", *Journal of Portfolio Management* 11, 9-17.

Ross, S.A. (1976): "The arbitrage theory of capital asset pricing", *Journal of Economic Theory* 13, 341–360.

Rossi, A.G. (2018): "Predicting stock market returns with machine learning", mimeo, University of Maryland.

Schaeffer, R., M. Khona, Z. Robertson, A. Boopathy, K. Pistunova, J.W. Rocks, I.R. Fiete

and O. Koyejo (2023): "Double descent demystified: identifying, interpreting and ablating the sources of a deep learning puzzle", https://arxiv.org/abs/2303.14151.

Sentana, E. (2005): "Least squares predictions and mean-variance analysis", *Journal of Financial Econometrics* 3, 56-78.

Sentana, E. (2009): "The econometrics of mean-variance efficiency tests: a survey", *Econometrics Journal* 12, C65-C101.

Sharpe, W.F. (1963): "A simplified model for portfolio analysis", *Management Science* 9, 277–293.

Sharpe, W.F. (1994): "The Sharpe ratio", *Journal of Portfolio Management* 21, 49-58.

Simon, F., S. Weibels and T. Zimmermann (2023): "Deep parametric portfolio policies", https://ssrn.com/abstract=4150292.

Spiegel, M. (2008): "Forecasting the equity premium: where we stand today", *Review of Financial Studies* 24, 1453-1454.

Stattman, D. (1980): "Book values and stock returns", *The Chicago MBA: A Journal of Selected Papers* 4, 25-45.

Tetlock, P. C. (2007): "Giving content to investor sentiment: the role of media in the stock market", *Journal of Finance* 62, 1139–1168.

Tobek, O. and M. Hronec (2021): "Does it pay to follow anomalies research? Machine learning approach with international evidence", *Journal of Financial Markets* 56.

Tobin, J. (1958): "Liquidity preference as behavior toward risk", *Review of Economic Studies* 25, 65-86.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society* B 58, 267-288.

Wang, H. and X.Y. Zhou (2020): "Continuous-time mean-variance portfolio selection: a reinforcement learning framework", *Mathematical Finance* 30, 1273–1308.

Yuan, M. and G. Zhou (2023): "Why naive 1/N diversification is not so naive, and how to beat it?", forthcoming in *Journal of Financial and Quantitative Analysis*.

Zhang, W. and S. Skiena (2010): "Trading strategies to exploit blog and news sentiment", Proceedings of the 4th international AAAI Conference on Weblogs and Social Media.