The Stochastic Discount Factor: Theory and Estimation

Empirical Asset Pricing

Mads Markvart Kjær

Department of Economics and Business Economics, Aarhus University, CREATES

E-mail: mads.markvart@econ.au.dk

Spring 2023

Why do we care?

Central asset pricing **questions**:

- Why do **different** assets give **different** (expected) **returns**?
- Is a **certain risk priced** in financial markets?
- How do we **interpret** compensation/**risk premia**?

➨ For trying to **answer** these questions, the dominant approach is to use the **stochastic discount factor**

- In other words: you will learn to analyse the existence and formulation of the SDF, how the existence relates to arbitrage, in addition to how it helps us pricing financial assets.

SDFs are everywhere!

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

ELSEVIER

Cross-sectional return dispersion and currency momentum☆

Jonas N. Eriksen

CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark

3.1. Methodology

In the absence of arbitrage, risk-adjusted currency excess returns have a price of zero and satisfy the basic Euler equation

$$E_t\left[M_{t+1} RX^j_{t+1}\right] = 0, \tag{3}$$

where RX^j_{t+1} is the excess return on currency portfolio j at time $t+1$ and M_{t+1} is a stochastic discount factor (SDF) that is linear in the risk factors f_{t+1}

$$M_{t+1} = 1 - b'\left(f_{t+1} - \mu_f\right), \tag{4}$$

where b is a vector of factor loadings and μ_f denotes factor means. This specification implies a beta pricing model

$$E\left[RX^j_{t+1}\right] = \lambda' \beta^j, \tag{5}$$

where the expected excess currency return depends on the factor risk prices λ and the corresponding factor betas β^j. The factor price

SDFs are everywhere!

The Journal of FINANCE

The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXVII, NO. 2 • APRIL 2012

Carry Trades and Global Foreign Exchange Volatility

LUKAS MENKHOFF, LUCIO SARNO, MAIK SCHMELING, and ANDREAS SCHRIMPF[*]

We denote excess returns of portfolio i in period $t+1$ by rx^i_{t+1}.[17] The usual no-arbitrage relation applies so that risk-adjusted currency excess returns have a price of zero and satisfy the basic Euler equation

$$\mathbb{E}\left[m_{t+1} rx^i_{t+1}\right] = 0, \tag{5}$$

with a linear SDF given by $m_t = 1 - b'(h_t - \mu)$, where h denotes a vector of risk factors, b is the vector of SDF parameters, and μ denotes the factor means. This specification implies a beta pricing model where expected excess returns depend on factor risk prices λ and risk quantities β_i, which are the regression betas of portfolio excess returns on the risk factors:

$$\mathbb{E}[rx^i] = \lambda' \beta_i, \tag{6}$$

SDFs are everywhere!

Currency Premia and Global Imbalances

Pasquale Della Corte
Imperial College London and Centre for Economic Policy Research

Steven J. Riddiough
University of Melbourne

Lucio Sarno
City University London and Centre for Economic Policy Research

4.1 Methodology

We denote the discrete excess returns on portfolio j in period t as RX_t^j. In the absence of arbitrage opportunities, risk-adjusted excess returns have a price of zero and satisfy the following Euler equation:

$$E_t[M_{t+1} RX_{t+1}^j] = 0, \tag{4}$$

with a stochastic discount factor (SDF) linear in the pricing factors f_{t+1}, given by

$$M_{t+1} = 1 - b'(f_{t+1} - \mu), \tag{5}$$

where b is the vector of factor loadings, and μ denotes the factor means. This specification implies a beta pricing model in which the expected excess return on portfolio j is equal to the factor risk price λ times the risk quantities β^j. The beta pricing model is defined as

Outcome of lecture

After the lecture, you should have

- **knowlegde** and **understading** of
  - The **stochastic discount factor** (SDF), its **existence** and **use in asset pricing**, its **implications** for **arbitrage** and market **completeness**, and its **relation** to the **investor's marginal utility**
- and be able to
  - **Discuss** and **estimate** the SDF using **common empirical methods**

➜ The questions you can (try to) answer using the SDF is not only exciting, but **the theory** is also **highly relevant** for the **exam**!

- First, we do need to distinguish between **complete** and **incomplete** markets

Complete markets

Complete markets, cont.

To explain the idea of **complete** markets, consider the **following setup**:

- **No transactions costs** and **perfect information** (no frictions)
- A **discrete-state** model with S many **states of the world**, $s = 1, \ldots, S$, each with **probability** $\pi(s)$
- For each **state** s, there exists a **contingent claim** that pays $**1** in state s and **nothing** in any other state (This is also known as an Arrow-Debreu asset)
- The **price** of the asset is $q(s)$

Complete markets

Properties of complete markets

- **All possible bets** of the **future states of the world** can be constructed using the contingent claims
- **Prices** on all contingent claims are **strictly positive**, $q(s) > 0$

- If $q(s) \leq 0$, we have an arbitrage oppotunity
  - Suppose $q(s) \leq 0$. The investors “buys" the asset for either **nothing** or even receives a **positive** payoff today and gets an asset that has
    1. **non-zero probability** for receiving a **positive payoff** if state s realizes in the next period,
    2. **zero probability** for a **negative payoff** in any future state

$\Rightarrow$ **infinitely attractive** investment

The fundamental equation of asset pricing

- The assets are only **distinguished** by their **state-dependent** payoffs $X(s)$, $s = 1, \ldots, S$
- Given the **finite state-space**, **all** assets can be **replicated** using **bundles of contingent claims**
- Under **no-arbitrage**, the **price of an asset** with payoff X is given as

$$p_i(X) = \sum_{s=1}^{S} q(s) X_i(s). \tag{1}$$

- Also known as **Cochrane's happy meal theorem**

The fundamental equation of asset pricing

Law of one price (intuitively)

The **law of one price** says, intuitively, that **two assets** with **identical** payoffs (characteristics) in every state **must have the same price**

- If this **does not hold**, it would imply **arbitrage opportunities**
- Why? Suppose the contrary, that is,

$$p_1 > p_2, \tag{2}$$

  but identical across all states.
- Buy asset 2, sell asset 1 yields:
  - $p_1 - p_2 > 0$ today
  - zero in next period with probability 1
- So a **violation of the law of one price** leads to **arbitrage**
- **Arbitrage** does, however, **not** necessarily lead to a **violation of the law of one price**

The fundamental equation of asset pricing

- To get an **expectational expression**, multiply (1) by $1 = \pi(s)/\pi(s)$

$$p(X) = \sum_{s=1}^{S} \pi(s) \frac{q(s)}{\pi(s)} X(s) = \sum_{s=1}^{S} \pi(s) M(s) X(s), \tag{3}$$

  where $M(s) = q(s)/\pi(s)$ is **defined** as the **SDF**
- **One-to-one mapping** between **SDF** and **risk-neutral probabilities**!

The fundamental equation of asset pricing

The **fundamental equation of asset pricing** reads

$$p(X) = \mathbb{E}[MX]. \tag{4}$$

Since $q(s), \pi(s) > 0$, it follows that $M(s) > 0$

- Let us put some economic intuition into the theory by an application of it ➨ **consumption-based asset pricing (CCAPM)**

The representative agent

- A **market equilibrium** consists of **many** (heterogeneous) **investors**, each optimizing their utility
- Wouldn't it be **nice** if we could **simplify the market** into a **single representative agent** and get the **same equilibrium**?

The aggregation property of the economy

- If **markets are complete**, financial markets have the **aggregation property**
- That is, **equilibrium prices** are the same as in a **hypothetical representative-agent economy**
- …, and we can **work** with a **single representative agent**

- Consumption-based asset pricing models frequently aggregate individual investors into a single utility-maximizing (representative) agent whose **utility** derives from **aggregate (per capita) consumption**

Consumption-based asset pricing framework

- Let $u(c_t)$ be the **concave**, **time-separable** **utility function**, where c denotes aggregate consumption (per capita)
- Each period, the investor **chooses** between **consumption** and **investing** (for future consumption) to optimally smooth consumption

The maximization problem

The **representative agent** **maximizes**

$$\max \sum_{t=1}^{T} \delta^t \mathbb{E}[u(c_t)|\mathcal{F}_t], \tag{5}$$

subject to budget constraints, which we leave unspecified for now, and a large T

- $\delta = (1+\tau)^{-1}$ is a (deterministic) **subjective discount factor**, and τ is the **subjective time preference rate**. The smaller δ, the more **impatient** the investor is ➨ it prefers consumption **now** versus in the **future**
- $\mathcal{F}_t$ is the time-t **filtration** (**information set** available to the investor).

Consumption-based asset pricing framework

Euler equation (version 1)

The **solution** to the **maximization problem** is

$$u'(c_t)P_{i,t} = \mathbb{E}\left[\delta u'(c_{t+1})(P_{i,t+1} + D_{i,t+1})|\mathcal{F}_t\right]. \tag{6}$$

where $P_{i,t}, D_{i,t}$ is the **price** and **dividend** paid by **asset** i at time t and u' is the first **derivative** of the **utility function** w.r.tc (marginal utility)

- The Euler equation is optimum for the representative investor's consumption and portfolio choice problem
- It **equates** marginal **cost** and **benefit** of current versus future consumption:
  1. LHS: **Marginal utility loss** in period t from buying one additional unit of the asset instead of consuming today
  2. RHS: **Expected discounted marginal utility gain** associated with buying an additional unit of the asset instead of consuming today

Consumption-based asset pricing framework

Euler equation (version 2)

The **solution** to the maximization problem can be **rewritten** to

$$P_{it} = \mathbb{E}\left[\delta \frac{u'(c_{t+1})}{u'(c_t)}(P_{i,t+1} + D_{i,t+1})|\mathcal{F}_t\right], \tag{7}$$

or, **equivalently**,

$$1 = \mathbb{E}\left[\delta \frac{u'(c_{t+1})}{u'(c_t)} R_{i,t+1}|\mathcal{F}_t\right], \tag{8}$$

where $R_{i,t+1} = (P_{i,t+1} + D_{i,t+1})/P_{i,t}$ is the (gross) return on asset i.

- Since we consider a period-by-period optimization problem, the payoff of asset i is $X_{i,t+1} = P_{i,t+1} + D_{i,t+1}$

Consumption-based asset pricing framework

- This **matches** the structure of the **simple fundamental equation** of asset pricing in (4)

Theorem: SDF in consumption-bassed asset pricing

In a **discrete-time**, **complete** market economy with a single consumption good, let δ be the time subjective discount factor and u the utility function of the representative individual with time-separable utility. Then (the process)

$$M_{t+1} = \delta \frac{u'(c_{t+1})}{u'(c_t)}, \quad t = 0,1,\ldots,T \tag{9}$$

is an **SDF** (at all time points).

Consumption-based asset pricing framework

- As such, the **price** of asset i at any time-point is the discounted value of the future payoff
- The discounting is the **marginal rate of substitution** between time t and $t+1$ consumption ➨ the growth in marginal utility.
- with a **large growth** in marginal utility, any future payoff is **highly valued** and the **price today** (expected return) will be **higher** (lower), and vice versa
- The **functional** form of M_{t+1} **depends** on the **choice of the utility function** and is extremely scrutinized in the academic literature
- …, for now, we will consider the most famous (and the most simple) example, i.e., power utility

Consumption-based asset pricing framework

- A compact notation is, thus,

$$1 = \mathbb{E}_t[M_{t+1}R_{i,t+1}], \tag{10}$$

using (9), and where we use subscript t to indicate **conditional moments**.

Central consumption-based asset pricing equation

The **central consumption-based asset pricing equation** is

$$\mathbb{E}_t[r_{i,t+1}] - r_{f,t+1} = -(1 + r_{f,t+1})\mathrm{Cov}_t[M_{t+1}, r_{i,t+1}], \tag{11}$$

where $r_{i,t+1}$ is the simple return, $r_{i,t+1} = R_{it+1} - 1$. Equivalently, using (9),

$$\mathbb{E}_t[r_{i,t+1}] - r_{f,t+1} = -\delta(1 + r_{f,t+1})\mathrm{Cov}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}, r_{i,t+1}\right]. \tag{12}$$

Consumption-based asset pricing framework

Consumption-based asset pricing logic

Assets with $\text{Cov}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}, r_{i,t+1}\right] < 0$ earn **higher expected excess returns**

- Note that $\frac{u'(c_{t+1})}{u'(c_t)}$ is inversely related to the business cycle:
    1. **high** during **recessions** (when consumption is low)
    2. **low** during **expansions** (when consumption is high)
- If $\text{Cov}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}, r_{i,t+1}\right] < 0$, asset i **pays off poorly in bad states** and well in good states, making it undesirable for consumption smoothing purpose
- If $\text{Cov}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}, r_{i,t+1}\right] > 0$, asset i **provides consumption insurance by paying off in bad states** when the investor values additional consumption most highly

Exchange rates as assets

- Remember from the SDF lecture for a foreign investor:

$$1 = \mathbb{E}_t(\tilde{R}_{t+1}\tilde{M}_{t+1}) \tag{13}$$

- For a domestic investor investing in the same asset, no-arbitrage also implies

$$1 = \mathbb{E}_t(\tilde{R}_{t+1}\frac{S_{t+1}}{S_t}M_{t+1}) \tag{14}$$

- Meaning that

$$\mathbb{E}_t(\tilde{R}_{t+1}\frac{S_{t+1}}{S_t}M_{t+1}) = \mathbb{E}_t(\tilde{R}_{t+1}\tilde{M}_{t+1}) \tag{15}$$

A suffient condition

- A sufficient condition for the eq. (15) is

$$\frac{S_{t+1}}{S_t} M_{t+1} = \tilde{M}_{t+1} \tag{16}$$

- Under no-arbitrage and completeness (uniqueness of the SDFs), the exchange rate is determined by

$$\frac{S_{t+1}}{S_t} = \frac{\tilde{M}_{t+1}}{M_{t+1}}, \tag{17}$$

where S_t is measured as foreign prices per unit of domestic prices (by far the worst part of working with currencies)

A suffient condition

- Compared to stocks, exchange rates have a tighter connection to interest rates, but are less correlated than ZCB with different maturities
- This implies, that if we are interested in examining, for instance, the impact of the macroeconomy on asset pricing, exchange rates are a natural asset to base your analysis on!

Incomplete markets

Incomplete markets

- What if markets are **incomplete**?
- Rather than **deriving a specific SDF** as in the consumption-based framework, we will now **work backward** (and be **more general**)
- Essentially, an **SDF** is just defined as the random variable that makes the following **representations true**

$$P_{i,t} = \mathbb{E}_t[M_{t+1} X_{i,t+1}] \quad \text{and} \quad 1 = \mathbb{E}_t[M_{t+1} R_{it+1}],\ \forall i,t$$

- When can we find such SDF, M_{t+1}?
- Can we **use** this representation **without implicitly assuming** all the structure of the investors, utility functions, complete markets, etc.?
- The **short answer** will be **yes**! …, under some conditions.

Incomplete markets

- **Suppose** we **observe** a set of asset payoffs X and prices P

Payoff space

The **payoff space**, denoted Ξ, is defined as the set of all the payoffs that investors can buy, including combining various assets

- To **obtain existence** of (at least one) **SDF**, we need to put some high-level structure on the economy

Incomplete markets

Assumptions

We make the following two assumptions:

1. **Portfolio formation**: $X_1, X_2 \in \Xi \Rightarrow X_p \equiv aX_1 + bX_2 \in \Xi$ for any real-valued a,b
2. **Law of one price**: $P(X_p) \equiv P(aX_1 + bX_2) = aP(X_1) + bP(X_2)$.

- Assumption 1 is quite restrictive in the sense that it rules out shorting constraints (by allowing a,$b < 0$), bid-ask spreads, leverage limitations, etc.
- …those can, however, be incorporated at the cost of complexity
- Assumption 2 is quite restrictive in the sense that it rules out the effect of packaging - a package is worth only what it contains and now how it is, e.g., branded - i.e. the happy meal theorem

Existence of an SDF

Theorem: Existence of an SDF

Given **portfolio formation** (Assumption 1) and the **law of one price** (Assumption 2), there **exists a payoff** $X^* \in \Xi$ such that

$$P(X) = \mathbb{E}[X^* X], \quad \forall X \in \Xi. \tag{18}$$

- That is, under **Assumption 1 and 2**, X^* **satisfies the fundamental equation of asset pricing**, (4), **without the positivity property,** $X^* > 0$**, ensured**
- As such, X^* is an SDF ➨ in (in)complete markets it is (not) unique.
- It also goes the other way, i.e., the **existence of an SDF** implies **Assumption 1 and Assumption 2**

Positivity of the SDF

- While **Assumption 1 and 2** ensure the **existence** of an SDF, it does **not guarantee positivity**
- Why do we need it to be positive? It naturally results from any sort of **utility maximization**
- Recall,

$$M_{t+1} = \delta \frac{u'(c_{t+1})}{u'(c_t)}. \tag{19}$$

- Since $\delta > 0$ and $u'(c) > 0$ (unreasonable to think that people will get more utility from consuming less), $M_{t+1} > 0$
- But positivity of the SDF also rules out negative prices for assets that pay positive payoffs

Positivity of the SDF

Absence of arbitrage

A payoff space Ξ and pricing function $P(X)$ have absence of arbitrage if every payoff with $X \geq 0$ with certainty and if every payoff with positive $X > 0$ with some positive probability has positive price $P(X) > 0$.

- This definition is slightly different from the one given in Campbell (2017) but it is more intuitive
- It means that you cannot get a portfolio for free that *might* pay off positively, but will never certainly cost you anything.

Positivity of the SDF

Theorem: Positivity and existence of the SDF

1. $P = \mathbb{E}[MX]$ and $M(s) > 0 \Rightarrow$ absence of arbitrage.
2. Absence of arbitrage $\Rightarrow \exists M$ such that $P = \mathbb{E}[MX]$ and $M(s) > 0$.

- That is, a **positive SDF exists** if and only if **markets are free of arbitrage**. If so, all assets can be priced according to the fundamental equation of asset pricing in (4)

Why does all this theory matter for an empirical exercise?

- In the end, every choice we make most be due to some maximization exercise
- A natural way to motivate a risk factor is how it relates to the maximization problem of the representative investor:

$$\max \sum_{t=1}^{T} \delta^t \mathbb{E}[u(c_t)|\mathcal{F}_t], \tag{20}$$

- If a **risk factor** has an impact on **risk aversion**, **consumption (opportunities)**, **investment opportunities**, or the time-separable **discount function**, it **directly affects** the **SDF** and, thereby, **expected returns**

SDF and β representation

SDF-talk: Properties

- The **fundamental pricing equation**, using the SDF, is one type of **representation of asset pricing**
- **Two others exist**; β representation and mean-variance frontier representation

"All are equivalent" representation theorem

…both representations are equivalent

1. SDF $\Rightarrow \beta$
2. $\beta \Rightarrow$ SDF

- That is, if an **SDF exists**, we can always **find a β representation for asset returns**, and vice versa
- Additional details can be found in Cochrane (2009) Ch. 6

SDF-talk: SDF $\Rightarrow \beta$

- Recall the **fundamental asset pricing equation** in returns

$$1 = \mathbb{E}_t[M_{t+1}R_{i,t+1}] \tag{21}$$

  expressed in a discrete-time multi-period fashion

- Recall that if (21) holds for all t, it must also hold unconditionally (use law of iterated expectations), such that

$$1 = \mathbb{E}[M_{t+1}R_{i,t+1}]. \tag{22}$$

- We will discuss the **unconditional** implications of the **conditional** models further later
- In the following, we will work with this unconditional implication ➨ it essentially puts focus on average returns

SDF-talk: SDF $\Rightarrow \beta$

Risk-free rate in SDF form

Consider the **risk-free asset**, denote it by "$i = f$", with payoff $X_{f,t+1} = 1$ in all states, with **certainty**. By the fundamental asset pricing equation we must then have

$$\mathbb{E}[P_{f,t}] = P_{f,t} = \mathbb{E}[M_{t+1}], \tag{23}$$

such that

$$R_{ft+1} = \mathbb{E}[M_{t+1}]^{-1}. \tag{24}$$

SDF-talk: SDF $\Rightarrow \beta$

- It follows by **general covariance rules** that

$$\begin{aligned} 1 &= \mathbb{E}[M_{t+1}R_{i,t+1}] \\ &= \mathbb{E}[M_{t+1}]\mathbb{E}[R_{i,t+1}] + \text{Cov}[M_{t+1},R_{i,t+1}]. \end{aligned} \tag{25}$$

- …such that

$$\mathbb{E}[R_{i,t+1} - R_{f,t+1}] = -R_{f,t+1}\text{Cov}[M_{t+1},R_{i,t+1}]. \tag{26}$$

- The return on any asset is:
  - The risk-free return
  - A term that informs about the **co-variation** between the **SDF** and **returns** (This is where all the intuition in asset pricing models comes from!)

SDF-talk: SDF $\Rightarrow \beta$

SDF $\Rightarrow \beta$ representation

It follows from (26) that (multiply by $\frac{\text{Var}[M_{t+1}]}{\text{Var}[M_{t+1}]}$)

$$\mathbb{E}[R_{i,t+1} - R_{f,t+1}] = \beta_{i,M}\gamma_M, \tag{27}$$

where

$$\beta_{i,M} = \frac{\text{Cov}[M_{t+1}, R_{i,t+1}]}{\text{Var}[M_{t+1}]} \tag{28}$$

is the (single) **regression coefficient** of any asset return R_{it+1} on the SDF, and

$$\gamma_M = -R_{f,t+1}\text{Var}[M_{t+1}] \tag{29}$$

is the **factor risk premium**, noting that $R_{f,t+1}$ is known with certainty. (*do not confuse the subscript M with "market" ➨ it is due to the SDF denoted by M*)

SDF-talk: SDF $\Rightarrow \beta$

- This relates directly to the **intuition** presented in the consumption-based framework ➨ **expected excess returns** are **linear** in the regression βs of asset returns on $M_{t+1} = (c_{t+1}/c_t)^{-\rho}$
- Typically, γ_M is treated as a **free parameter** and **estimated** in empirical evaluations of factor models, however according to theory it should equal $-R_{ft+1}\text{Var}[(c_{t+1}/c_t)^{-\rho}] < 0$

β representation implications

- For a choice/model of SDF, a β representation is thus implied (and can be estimated)
- **Differences in expected excess returns** among a cross-section of assets must be explained by **differences in their βs (risks)**
- This **defines** the **empirical approaches** to **estimation of asset pricing models** which we will see/cover in the next lecture

SDF-talk: $\beta \Rightarrow$ SDF

- **Suppose** we have an **expected return model** in β representation (for instance, the CAPM). What SDF does this imply?

$\beta \Rightarrow$ SDF representation

A β representation of expected returns are equivalent to linear models for the SDF as per

$$M_{t+1} = a - b' f_{t+1}, \tag{30}$$

where a,b are parameters and f_{t+1} the risk factors. We use negative b for expositional reasons, see e.g. the example with CCAPM below in (34)

SDF-talk: $\beta \Rightarrow$ SDF

- Typically, we make **two convenient assumptions** that are **without loss of generality**:

Assumptions (w.l.o.g.)

1. **De-meaned factors**: We assume that factors are de-meaned such that $\mathbb{E}[f_{t+1}] = 0$. This implies that $\mathbb{E}[M_{t+1}] = \mathbb{E}[a - b' f_{t+1}] = a$.
2. **Normalization**: We normalize the mean of the SDF to unity, i.e. $\mathbb{E}[M_{t+1}] = 1$, which under Assumption 1 just above implies that $a = 1$

- Note, we are only able to **identify the SDF** up to the **scale of a constant** since $M_{t+1} = a(1 - (b/a)' f_{t+1})$

SDF-talk: $\beta \Rightarrow$ SDF

$\beta \Rightarrow$ SDF representation theorem

Suppose Assumptions 1 (de-meaned factors) and 2 (normalization) holds. Given the following β representation,

$$\mathbb{E}[R_{it+1} - R_{ft+1}] = \beta_i' \gamma, \tag{31}$$

where β are multiple regression coefficients of excess returns on the factors, we can always find b such that

$$M_{t+1} = 1 - b' f_{t+1} \tag{32}$$

with $\mathbb{E}[M_{t+1}(R_{it+1} - R_{ft+1})] = 0$.

- Also, given (32), we can always find a γ such that (31) holds

SDF-talk: Interpretation

- From **the β representation**, it is clear that γ may be interpreted as the **price of the factor risk**, or the **factor risk premium**
- For every unit β, the expected excess return increases by γ

The fundamental research question

As such, a test of $\gamma \neq 0$ is often called **a test** of whether **the factor is "priced" in the financial markets**. This is typically the main research question posed in studies in empirical asset pricing

SDF-talk: Interpretation via CCAPM

Example of CCAPM

- The CCAPM (approximately) stipulates that there exist a single risk factor, which is the logarithmic growth rate in aggregate consumption, denoted by $f_{t+1} = \tilde{c}_{t+1}$
- Suppose it is de-meaned and that Assumption 2 is invoked, such that

$$M_{t+1} = 1 - b\tilde{c}_{t+1}. \tag{33}$$

- It then follows from (26) that

$$\begin{aligned} \mathbb{E}[R_{i,t+1} - R_{f,t+1}] &= -R_{f,t+1}\text{Cov}[M_{t+1}, R_{i,t+1}] \\ &= -R_{f,t+1}\text{Cov}[1 - b\tilde{c}_{t+1}, R_{i,t+1}] \\ &= R_{f,t+1}b\text{Cov}[\tilde{c}_{t+1}, R_{i,t+1}] \\ &= \beta_i^c \gamma^c, \end{aligned} \tag{34}$$

where $\beta_i^c = \text{Cov}[\tilde{c}_{t+1}, R_{i,t+1}]/\text{Var}[\tilde{c}_{t+1}]$ and $\gamma^c = R_{f,t+1}b\text{Var}[\tilde{c}_{t+1}]$

SDF-talk: Interpretation via CCAPM

Example of CCAPM

- If $b > 0$ higher consumption growth reduces marginal utility growth. In this case, we have that $\gamma^c > 0$ (assuming $R_{ft+1} > 0$)
- That is, $\beta_i^c > 0$ is compensated/priced in the financial markets

SDF-talk: Uniformity and challenges

- So, **an asset pricing framework** that initially seemed to require a **lot of structure** (the representative, utility-maximising agent in the consumption-based framework) turns out to require **minimal structure**
- **Under few appropriate assumptions**, we can **always** start an analysis by writing $P_{it} = \mathbb{E}_t[M_{t+1}X_{t+1}]$ or $0 = \mathbb{E}_t[M_{t+1}(R_{it+1} - R_{ft+1})]$, **for any asset** (equity, bond, currency (we will later return to the asset class and how it relates to SDFs), house, cryptocurrency, you name it)

SDF-talk: Uniformity and challenges

- …and **this does not require** any **assumptions** on **market completeness**, **contingent-claim** or **representative agent**
- Of course, **this it not without a cost**: **all the economic, statistical, and predictive content** comes from **picking the SDF** model, i.e. $M_{t+1} = h(\text{data}_{t+1}, \theta)$, for some function $h(\cdot, \theta)$

GMM

Generalized Methods of Moments (GMM)

- **Generalized Methods of Moments** (GMM) is an **estimation** principle, using **moment conditions** to enable identification.
- **Nests OLS**, **instrumental variables**, and **MLE**
- **Moment conditions** are of the form

$$\mathbb{E}[G(\text{data}_t,\theta)] = 0, \tag{35}$$

where $G(\cdot)$ is a N-dimensional **function of data** and a K-dimensional vector of **parameters**, θ, that is to be estimated.

Motivation

- **Robust** to assumptions about **homoskedasticity and autocorrelation**
- **Robust** to **distributional assumptions**
- **Can handle non-linear models**
- **Can handle economic models** that are formulated directly as **moment conditions**
- …and it is **extremely useful** for estimating (linear) β **represented factor models**, taking into account important statistical/empirical issues such as **errors-in-variables, autocorrelation, and heteroscedasticity**

Estimation approach

- To **estimate model parameters**, we consider the **sample average counterpart** of the moment conditions, called the **object function**:

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} G(\text{data}_t, \theta), \tag{36}$$

  where T is the sample time series dimension

- Note that $g_T(\theta) \xrightarrow{p} \mathbb{E}[G(\text{data}_t, \theta)]$ as $T \to \infty$.

Estimation principle (intuition)

GMM estimates parameters as those that make the **object function**, $g_T(\theta)$, **as close** to the ones implied by the **moment conditions**, i.e., 0.

Estimation approach

- We **need at least as many moment conditions** as we have **model parameters**, $N \geq K$:
  1. If $N < K$, the model is **not identified**
  2. If $N = K$, the model is **exactly identified**, sometimes with an analytical solutions if $G(\cdot)$ is linear in θ
  3. If $N > K$, the system is **overidentified** and **numerical optimization** is needed
- We need a way to **weight** each moment condition in the estimation, denoting this **weighting matrix** by A_T

Estimation approach

Estimation principle (formally)

For **a given choice** of **weighting matrix** (discussed below), GMM estimates parameters by **minimizing** a quadratic form of the weighted sample moment condition as per

$$\hat{\theta} = argmin_{\theta} g_T(\theta)' A_T g_T(\theta) \tag{37}$$

- For **any choice of weighting matrix** (e.g. the identity matrix), the **GMM estimator is consistent**, $\hat{\theta} \xrightarrow{p} \theta$ as $T \to \infty$
- The estimation procedure is often done in two or several steps, coined **two-stage** and **iterated** GMM
- To understand why, we need to understand the choice of weighting matrix

Choice of weighting matrix

- In the case of exact identification, we have that all moment conditions can be set equal to zero
- In the overidentified case, this is no longer possible
- The **weighting matrix** **determines the weight** each moment should have when estimating the parameters ➨ a very **important choice**

Symmetric (or equal) weights

If the weighting matrix is set to the identity matrix, it puts equal emphasis on all moment conditions, that is,

$$A_T = I_N,$$

where I_N is the $N \times N$-dimensional identity matrix

Choice of weighting matrix

- One particular choice of A_T is **optimal** in a **statistical sense**
- …in the sense that the resulting GMM estimator has the **lowest asymptotic covariance matrix** among all possible GMM estimators

Optimal weights

If the **weighting matrix** is set to the **inverse of the long-run covariance matrix**, it puts most weight on the sample moments with lowest sampling variation, that is,

$$A_T = S^{-1},$$

where S is the long-run covariance matrix of the sample moments defined on the following slide

- Suppose moment conditions are asset pricing errors. Then this weighting matrix puts **most (least) weight** on the assets with **least (most) variance** of their pricing errors

Choice of weighting matrix

- The **long-run covariance matrix** is defined as

$$
\begin{aligned}
S &\equiv \lim_{T\to\infty} \mathrm{Var}[\sqrt{T} g_T(\theta_0)] \\
&= \sum_{s=-\infty}^{\infty} \mathbb{E}[G(\mathrm{data}_t, \theta_0) G(\mathrm{data}_{t-s}, \theta_0)'] \qquad (38)
\end{aligned}
$$

  where θ_0 is the population (true) parameters.

- If observations are independent, this reduces to

$$
S = \mathbb{E}[G(\mathrm{data}_t, \theta_0) G(\mathrm{data}_t, \theta_0)']. \qquad (39)
$$

- The estimator of S, $\hat{S}$, **requires estimated parameters**, $\hat{\theta}$, and is, as such, infeasible at first (put “hats" on everything unknown in the equations) …for that reason, **we need an additional step**

Two-stage and iterated GMM

Two-stage GMM

1. **Estimate GMM parameters**, using (37), with A_T equal to an arbitrary, but fixed, choice of matrix. Often, this is $A_T = I_N$. This generates $\hat{\theta}^{(1)}$, which is consistent and asymptotically normal. Use $\hat{\theta}^{(1)}$ to estimate $\hat{S}$
2. **Estimate second-stage GMM parameters**, using (37), with $A_T = \hat{S}^{-1}$. This generates $\hat{\theta}^{(2)}$, which is consistent and asymptotically normal. (Note that there is an error in Campbell (2017), as he forgets to invert $\hat{S}$ in his equation (4.88).)

- Note that the asymptotic properties are similar for each stage, yet in finite samples it is sometimes **beneficial to continue the procedure** by using $\hat{\theta}^{(2)}$ to update the estimate of $\hat{S}$ and then re-estimate parameters to get $\hat{\theta}^{(3)}, \ldots,$ until one stops when the errors, $Q(\hat{\theta}) = g_T(\hat{\theta})' A_T g_T(\hat{\theta})$, are sufficiently small

Asymptotic distribution and hypothesis testing

Asymptotic distributions for arbitrary weighting matrix

As $T \to \infty$ and any fixed A_T, it holds that

$$\hat{\theta} \xrightarrow{d} N(\theta_0, T^{-1}V), g_T(\hat{\theta}) \xrightarrow{d} N(0, T^{-1}\Omega),$$

where "$\xrightarrow{d}$" means **convergence in distribution**

$$V = (D'A_T D)^{-1} D' A_T S A_T (D' A_T D)^{-1}, \tag{40}$$

$$\Omega = \left(I_N - D(D'A_T D)^{-1} D' A_T\right) S \left(I_N - A_T D (D' A_T D)^{-1} D'\right)', \tag{41}$$

and D is $\mathbb{E}[\partial g_T(\theta_0)/\partial\theta_0]$ the gradient. For $A_T = S^{-1}$ the above simplifies to:

$$V = \left(D' S^{-1} D\right)^{-1}, \tag{42}$$

$$\Omega = \left(S - D(D'S^{-1}D)^{-1} D'\right). \tag{43}$$

Hansen's J-test for overall fit

- As a test of the **overall fit** of the model, one may apply **Hansen's J-test** (also known as a test for overidentifiying restions)
- This test examines whether $g_T(\hat{\theta})$ is sufficiently close to zero

Hansen's J-test for overall fit

For the **arbitrary weighting matrix**, Hansen's J-test is defined as

$$J_T \equiv g_T(\hat{\theta})'\hat{\Omega}^+ g_T(\hat{\theta}) \xrightarrow{d} \chi^2_{N-K}, \tag{44}$$

where $\hat{\Omega}^+$ denotes the (Moore-Penrose) pseudoinverse of the (estimated) sample moment covariance matrix. (Error in Campbell (2017) eq. (4.82), missing a transpose in the first term.)

- Note that Campbell (2017) applies a quite different procedure in estimation and testing the overidentifiying restrictions

Inference on parameter(s)

- We can also **make hypothesis tests** on whether a **parameter** (or a group of parameters) is **equal to zero** (or something else for that matter)
- For a single, the i'th, parameter, we form a conventional t-statistic as per

$$\frac{\hat{\theta}_i}{\sqrt{\text{Var}[\hat{\theta}_{ii}]}} \xrightarrow{d} N(0,1), \tag{45}$$

  where $\text{Var}[\hat{\theta}_{ii}]$ is the i'th diagonal element of the estimate of the covariance matrix of parameters, $\hat{V}$.
- **For a group** of p many parameters, we form a **conventional Wald-type statistic** as per

$$\hat{\theta}_j' \text{Var}[\hat{\theta}_{jj}]^{-1} \hat{\theta}_j \xrightarrow{d} \chi^2_p, \tag{46}$$

  where $\hat{\theta}_j$ is a subvector of parameters and $\text{Var}[\hat{\theta}_{jj}]$ a submatrix of $\hat{V}$

Estimation of covariance matrix and D

- Regardless of the choice of weighting matrix in the estimation, to make inference we need:

  1. An **estimator of the long-run covariance**
  2. An estimate for D

- If the derivative is not easily obtainable in analytical form (which it is in many cases later in our lecture), numerical differentiation is easier

Estimation of gradient D

- To see the intuition, suppose θ is one-dimensional. The one-sided or forward numerical derivative is then

$$\hat{D} = \frac{g_T(\hat{\theta} + h) - g_T(\hat{\theta})}{h},$$

  where h is a very small number, e.g. $h = 1\text{e-}6$
- This is motivated from the definition of a derivative by

$$\lim_{\varepsilon \to 0} \frac{g_T(\hat{\theta} + \varepsilon) - g_T(\hat{\theta})}{\varepsilon}.$$

- We use this forward version for computational reasons
- If θ is multi-dimensional, one needs the **gradient**, and the numerical differentiation is conducted with respect to each element in θ.

Estimation of covariance matrix

- When estimating the **long-run covariance matrix**, S, we will distinguish between cases with or without **serial correlation**
- The first case without serial correlation can actually be motivated from the asset pricing context (see below)

Long-run covariance matrix estimation

Under **no serial correlation**, the long-run covariance matrix, S, is estimated by

$$\hat{S}(\hat{\theta}) = T^{-1} \sum_{t=1}^{T} \left(G(\text{data}_t, \hat{\theta}) - \bar{G}(\text{data}_t, \hat{\theta}) \right) \times \left(G(\text{data}_t, \hat{\theta}) - \bar{G}(\text{data}_t, \hat{\theta}) \right)', \tag{47}$$

where $\bar{G}(\text{data}_t, \hat{\theta}) = T^{-1} \sum_{t=1}^{T} G(\text{data}_t, \hat{\theta}) = g_T(\hat{\theta})$.

Estimation of covariance matrix

- If theory does not imply no serial correlation (see below), or if we want to construct tests that are robust to the presence of serial correlation, we have a **parametric** or **nonparametric** approach
- The **parametric** approach estimates a **VARMA model** for $G(\text{data}_t, \theta)$
- Alternatively, we can estimate S **nonparametrically** by a **heteroskedasticity- and autocorrelation-consistent** (HAC) covariance matrix estimator
- This is essentially a **weighted average** of all sample autocovariances of $G(\text{data}_t, \hat{\theta})$

Estimation of covariance matrix

HAC estimator of long-run covariance matrix

HAC estimators of the long-run covariance matrix take the form

$$\hat{S}_{HAC}(\hat{\theta}) = \hat{\Gamma}_0 + \sum_{i=1}^{T-1} \omega_i \left(\hat{\Gamma}_i + \hat{\Gamma}_i'\right), \tag{48}$$

where ω_i is a kernel (or weight), and

$$\begin{aligned}\hat{\Gamma}_i = T^{-1} \sum_{t=i+1}^{T} & \left(G(\text{data}_t,\hat{\theta}) - \bar{G}(\text{data}_t,\hat{\theta})\right) \\ & \times \left(G(\text{data}_{t-i},\hat{\theta}) - \bar{G}(\text{data}_{t-i}, \hat{\theta})\right)' \end{aligned} \tag{49}$$

is the i'th sample autocovariance matrix

Estimation of covariance matrix

- Higher-order autocovariances need to be down-weighted to ensure consistency and positive semi-definiteness in all (finite) samples
- A common kernel choice is the Bartlett kernel by Newey and West (1987), given by

$$\omega_i = \begin{cases} 1 - \frac{i}{m+1}, & \text{for } i \leq m+1, \\ 0, & \text{for } i \geq m+1, \end{cases}$$

  where $m \geq 0$, $m \in \mathbb{Z}$, is the bandwidth that controls the number of autocovariances included in the estimator.
- In practice, one needs to make sure that the choice of m does not leave out important autocovariances
- …by, e.g., trying different candidate values and ensuring that adding additional autocovariances will not affect the HAC estimate significantly (or picking it optimally and data-driven (Andrews, 1991))

Asset pricing meets GMM

Asset pricing meeting GMM

- While the **moment condition**s in GMM are all **unconditional** in the presentation so far, most **asset pricing models** imply results for **conditional moments** (e.g. the CCAPM), as per

$$\mathbb{E}[G(\text{data}_t, \theta) | \mathcal{F}_t] = 0. \tag{50}$$

- This essentially requires explicit modelling of the conditional distributions, which is often complicated
- Rather, we can focus on the **implications** for **unconditional models** derived from **conditional models** and test those

Asset pricing meeting GMM

- An **asset pricing model** expressed in **conditional moments** implies **two sets of unconditional moment constraints**:
  - A conditioning down principle
  - Instruments that stand in for conditioning information in $\mathcal{F}_t$

Asset pricing meeting GMM

Implication 1: Conditioning down

Taking **unconditional expectations** of (21) and using the **law of iterated expectations** yields

$$\begin{aligned}\mathbb{E}[P_{it}] &= \mathbb{E}\left[\mathbb{E}_t[M_{t+1}X_{it+1}]\right] \\ &= \mathbb{E}[M_{t+1}X_{it+1}]. \end{aligned} \tag{51}$$

- This has a similar structure as the conditional expression, yet the implied moment condition is

$$\mathbb{E}[M_{t+1}X_{it+1} - P_{it}] = 0, \tag{52}$$

with $G(\cdot) = M_{t+1}X_{it+1} - P_{it}$.

Asset pricing meeting GMM

- Let z_t be a so-called **instrument** observed at time t. For any random variable y_{t+1} it can be shown that, if

$$\mathbb{E}[y_{t+1}z_t] = 0, \quad \forall z_t \in \mathcal{F}_t, \tag{53}$$

  then it implies

$$\mathbb{E}[y_{t+1}|\mathcal{F}_t] = 0. \tag{54}$$

- Setting $y_{t+1} = M_{t+1}X_{it+1} - P_{it}$ reveals that

$$\mathbb{E}[(M_{t+1}X_{it+1} - P_{it})z_t] = 0, \quad \forall z_t \in \mathcal{F}_t, \tag{55}$$

  is **sufficient** for **estimating/testing the conditional model** of $P_{it} = \mathbb{E}_t[M_{t+1}X_{it+1}]$

Asset pricing meeting GMM

- Start with the **fundamental pricing equation** in (21) and multiply an instrument to get

$$P_{it} z_t = \mathbb{E}_t [M_{t+1} X_{it+1} z_t], \tag{56}$$

where z_t can “move freely" in and out of expectations as it is adapted to $\mathcal{F}_t$ (known at time t).

Implication 2: Scaled payoffs

Unconditional expectations and the law of iterated expectations yield

$$\mathbb{E}[P_{it} z_t] = \mathbb{E}[M_{t+1} X_{it+1} z_t]. \tag{57}$$

- Doing this for all z_t **generates a set of implications** **not captured by Implication 1**

Asset pricing meeting GMM

- The moment conditions implied are thus

$$\mathbb{E}[(M_{t+1}X_{it+1} - P_{it})z_t] = 0, \tag{58}$$

  where $G(\cdot) = (M_{t+1}X_{it+1} - P_{it})z_t$.
- In practice, we of course have to choose a limited set of instruments ➨ natural source of critique.
- It can be understood in the context of scaled payoffs and managed portfolios
- $\tilde{X}_{it+1} = X_{it+1}z_t$ is an alternative asset with a scaled payoff and it has price $\tilde{P}_{it} = P_{it}z_t$. Here, z_t is a weighting variable, that informs the manager/investor on how much to buy or sell of a given asset
- For instance, high z_t can be informative/forecast high returns and he/she should buy more and vice versa. As such, z_t scales the investment, naturally scaling the payoff and the price

Asset pricing meeting GMM

- Using Implication 1 and 2 is, in principle, sufficient for capturing all unconditional implications of the conditional model

Implications 1 and 2 in return form

We will almost always work with returns (to ensure stationary data) and the resulting moment conditions are for each return

$$\text{Implication } 1 : \mathbb{E}[M_{t+1}R_{it+1} - 1] = 0, \tag{59}$$

$$\text{Implication } 2 : \mathbb{E}[(M_{t+1}R_{it+1} - 1)z_t] = 0, \quad \forall z_t \in \mathcal{F}_t. \tag{60}$$

- Suppose we have several, e.g. n many, tests asset ➡ we then have a vector returns
- Denote this by $R_{t+1} = (R_{1t+1}, \ldots, R_{nt+1})'$.

Asset pricing meeting GMM

- Moreover, suppose we have q many instruments (excluding the constant), which we gather in a vector $Z_t = (1, z_{1t}, \ldots, z_{q,t})'$ (that includes the constant)
- We can then **express both implications** compactly in a **single equation**

Kronecker formulation of Implications 1 and 2

Implications 1 and 2, using R_{t+1} and Z_t, reads

$$\mathbb{E}[(M_{t+1}R_{it+1} - 1) \otimes Z_t] = 0, \tag{61}$$

where "$\otimes$" is the **Kronecker/tensor product** and means "multiply every element by every other element". This leads to $n(q+1)$ moment conditions

Asset pricing meeting GMM

Example: Kronecker formulation

Suppose $Z_t = (1, z_{1t})'$ and $R_{t+1} = (R_{1t+1}, R_{2t+1})'$. Then (61) is

$$\mathbb{E}\left[\begin{pmatrix} M_{t+1}R_{1t+1} \\ M_{t+1}R_{2t+1} \\ M_{t+1}R_{1t+1}z_{1t} \\ M_{t+1}R_{2t+1}z_{1t} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ z_{1t} \\ z_{1t} \end{pmatrix}\right] = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{62}$$

yielding $n(q+1) = 2(1+1) = 4$ moment conditions

Asset pricing meeting GMM: Summary

- The **asset pricing model** says that although conditional expected returns can vary over time, **discounted returns** should **always be the same, 1**
- The model prediction error is $U_{it+1} \equiv M_{t+1}R_{it+1} - 1$. The asset pricing model says it should be both conditionally and unconditionally zero (Implication 1)
- If the asset pricing model is supposed to be **true**, we should not be able to use **any information today**, i.e., z_t, to **forecast any of the errors** (Implication 2). This is exactly what testing $\mathbb{E}[U_{it+1}z_t] = 0$ means

Asset pricing meeting GMM

No serial correlation as per asset pricing models

Recall that the asset pricing models delineate that $\mathbb{E}[U_{t+1}z_t] = 0$ and $\mathbb{E}[U_{t+1}] = 0$. This also means that these conditions imply that if $z_t = U_t$, it has to satisfy $\mathbb{E}[U_{t+1}U_t] = 0$. That is, **they imply no serial correlation** since

$$\text{Cov}[U_{t+1}, U_t] = \mathbb{E}[U_{t+1}U_t] - \mathbb{E}[U_{t+1}]\mathbb{E}[U_t] = \mathbb{E}[U_{t+1}U_t] = 0.$$

- ...but **use the HAC covariance matrix anyway**, since no asset pricing model is really the true one!

Choosing weighting matrix A_T

- Recall that the **choice of weighting matrix** is essentially a choice on **how to weight the sample moments** in the GMM estimation
- This mostly comes down to choosing between setting $A_T = I_N$ or $A_T = S^{-1}$
- In the **context of asset pricing**, the **former weights all pricing errors equally** among assets, whereas the latter puts more emphasis on those **assets that are most precisely predicted**

Choosing weighting matrix A_T

1. If a **single model** is estimated and inference on its asset pricing ability is made only on this model, it is recommended to use the optimal weighting matrix $A_T = S^{-1}$
2. If a **pair or several models** are estimated and their asset pricing abilities compared, it is recommend to use the identity weighting matrix $A_T = I_N$

Choosing weighting matrix A_T

- Since S^{-1} (most likely) changes according to the model, one model may "improve" $J_T \equiv g_T(\hat{\theta})'\hat{\Omega}^{+} g_T(\hat{\theta})$ simply because it blows up S rather than making the pricing errors smaller
- Moreover, **if the risk-free rate** is included as a test asset (which it typically is), then J_T essentially evaluates how well each model prices the Risk-free bond if S^{-1} is used, **ignoring all the other assets**
- As such, **one has to use a common weighting matrix across all models** to answer whether one model leads to s**maller pricing errors** (describes data better) than others

A last comment

Can we put even less structure on the model?

- Take a look at Equation (26):

$$\mathbb{E}[R_{i,t+1} - R_{f,t+1}] = -R_{f,t+1}\text{Cov}[M_{t+1}, R_{i,t+1}]. \tag{63}$$

- Rewritting the equation yields:

$$\frac{\mathbb{E}[R_{i,t+1} - R_{f,t+1}]}{\sigma_i} = -R_{f,t+1}\sigma_M\text{corr}[M_{t+1}, R_{i,t+1}]. \tag{64}$$

- The maximum Sharpe ratio portfolio has a correlation of -1 with sdf $\rightarrow$ The maximum Sharpe ratio portfolio can be viewed as a conditional projection of the true SDF!
- ... and the SDF is mean-variance efficient

The unconditional mean-variance efficient portfolio

- We will now consider the framework of Chernov et al. (2022)
- Consider the UMVE weights and returns:

$$\omega_t^* = \frac{1}{1 + \mu_t' \Omega^{-1} \mu_t} \Omega^{-1} \mu_t \tag{65}$$

$$R_{t+1}^* = \omega_t *' R_{t+1}^e \tag{66}$$

- Define the SDF as

$$M_{t+1}^* = 1 - (R_{t+1}^* - E(R_{t+1}^*)) \tag{67}$$

- satisfies for all admissible portfolios

$$\mathbb{E}_t(M_{t+1}^* R_{t+1}^e) = 0 \tag{68}$$

UMVE implementation

- All we need to proxy the SDF is return and covariance matrix forecasts!
- ... And we have absolutely no structure on expected returns and the covariance matrix. We could apply simple linear models, factor models, ML, etc.
- ... But the universe of stocks contains roughly 5000 stocks making the covariance $\rightarrow$ extremely difficult to estimate without any structure...
- Therfore, Chernov et al. (2022) consider currencies, while Randl et al. (2022) apply the framework on international government bonds
- During the course, we will examine different ways to construct return forecasts, while we will not dig deeper into the covariance matrix. Chernov et al. (2022) applies a specific shrinkage method to estimate Σ. The code is uploaded on brightspace

Implementation

- Live scripts to estimate a simple CCAPM model using GMM and implementing the UVME for currencies are on brightspace

References

ANDREWS, D. W. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 817–858.

CAMPBELL, J. Y. (2017): *Financial decisions and markets: a course in asset pricing*, Princeton University Press.

CHERNOV, M., M. DAHLQUIST, AND L. A. LOCHSTOER (2022): "Pricing currency risks," *Journal of Finance*.

COCHRANE, J. H. (2009): *Asset pricing: Revised edition*, Princeton university press.

NEWEY, W. K. AND K. D. WEST (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

RANDL, O., G. SIMION, AND J. ZECHNER (2022): "Pricing and Constructing International Government Bond Portfolios," *Available at SSRN 4021429*.

Cross-sectional Asset Pricing

Empirical Asset Pricing

Mads Markvart Kjær

Department of Economics and Business Economics, Aarhus University, CREATES

E-mail: mads.markvart@econ.au.dk

Spring 2023

Outcome of lecture

After the lecture, you should have

- knowlegde and understading of
  - Econometric methods and techniques for estimating and testing risk premia in the cross- section of asset returns
- and be able to
  - Discuss and estimate the SDF using common empirical methods, evaluate its ability to price the cross-section and conduct valid inference, and reflect on the findings and their implications

➜ The methodology to conduct a cross-sectional empirical asset pricing study!

Objective of today

- In other words:

Elephants and the Cross-Section of Expected Returns

Nora Laurinaityte[§] Christoph Meinerding[†] Christian Schlag[‡] Julian Thimme[*]

➜ You should be able to test whether any given factor is priced in a cross-section of assets!

The main idea of empirical asset pricing

- Recall from the last lecture that we have equivalence between SDF and β representations
- …given an SDF, we can always find a β representation and given a β representation, we can always find a linear factor model that defines the SDF
- Linear factor models of the SDF and, equivalently, β representations are by far the most popular in the empirical asset pricing literature (as you will see):

The main idea of empirical asset pricing

Journal of Financial Economics 142 (2021) 1017–1037

ELSEVIER

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Is there a risk-return tradeoff in the corporate bond market? Time-series and cross-sectional evidence☆

Jennie Bai[a], Turan G. Bali[b,*], Quan Wen[b]

[a] McDonough School of Business, Georgetown University, and NBER Research Associate, 3700 O St., Washington, DC 20057, United States
[b] McDonough School of Business, Georgetown University, 3700 O St., Washington, DC 20057, United States

the factor model of Bai et al. (2019) that introduces the downside risk, credit risk, and liquidity risk factors based on independently sorted bivariate portfolios of bond-level credit rating, value at risk (VaR), and illiquidity:[14]

$$R_{i,t} = \alpha_i + \beta_{1,i} \cdot MKT_t + \beta_{2,i} \cdot DRF_t + \beta_{3,i} \cdot CRF_t + \beta_{4,i} \cdot LRF_t + \epsilon_{i,t}, \tag{10}$$

where $R_{i,t}$ is the excess return on bond i in month t. Total risk of bond i is measured by the variance of $R_{i,t}$ in Eq. (9), denoted by σ_i^2. Idiosyncratic (or residual) risk of bond i is

The main idea of empirical asset pricing

Common Risk Factors in Currency Markets

Hanno Lustig
UCLA Anderson and NBER

Nikolai Roussanov
Wharton, University of Pennsylvania and NBER

Adrien Verdelhan
MIT Sloan and NBER

$$M_{t+1} = 1 - b(\Phi_{t+1} - \mu_{\Phi}),$$

where b is the vector of factor loadings and μ_{Φ} denotes the factor means. This linear factor model implies a beta pricing model: The expected excess return is equal to the factor price λ times the beta of each portfolio β^j:

$$E[Rx^j] = \lambda' \beta^j,$$

The main idea of empirical asset pricing

COMMON RISK FACTORS IN CRYPTOCURRENCY

Yukun Liu
Aleh Tsyvinski
Xi Wu

Table 9: Cryptocurrency Market and Size Factor Model

$$R_i - R_f = \alpha^i + \beta^i_{CMKT} CMKT + \beta^i_{CSMB} CSMB + \epsilon_i \quad (2)$$

where $CMKT$ is the cryptocurrency excess market returns and $CSMB$ is the cryptocurrency size factor.

The main idea of empirical asset pricing

Journal of Banking & Finance 40 (2014) 346–363

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

ELSEVIER

Are there common factors in individual commodity futures returns?

CrossMark

Charoula Daskalaki [a], Alexandros Kostakis [b], George Skiadopoulos [a,c,*]

[a] Department of Banking and Financial Management, University of Piraeus, Greece
[b] Manchester Business School, University of Manchester, UK
[c] School of Economics and Finance, Queen Mary, University of London, UK

3. Asset pricing models: Macro and equity-motivated tradable factors

In this section, we investigate whether models that include aggregate and equity-motivated tradable factors can explain the common variation of commodity futures returns. Let the beta formulation of a K-factor asset pricing model

$$E(r_i) = \beta_i'\lambda, \quad i = 1, 2, \dots, N, \qquad (1)$$

The main idea of empirical asset pricing

The main idea

The main approach for testing asset pricing models is the use of time-series regressions and cross-sectional regressions. Either:

- Conducted separately as in Fama-Macbeth two-stage regressions
- Jointly in a unified GMM framework.

The main idea of empirical asset pricing

- Either one postulates a β representation of a model (e.g. derived from CAPM) or one starts with a linear factor model for the SDF (often the case with CCAPM).
- In the latter case, one can estimate the SDF loadings (b) directly and the back out the risk premia (γ) in the β representation directly.
- Otherwise, one can always work with the (possibly implied) β representation.

  $\rightarrow$We will focus on this approach today and see an example with the SDF being starting point later.

Which factors to use?

- In empirical asset pricing an essential problem is the choice of factors (more about this in next (or end of the this) week).

Factor selection

There are, broadly speaking, three common ways to select factors for the asset pricing model:

1. Theoretical or economic intuition:
    - Factors can be directly derived in theoretical asset pricing models like CAPM or CCAPM
    - They can be motivated via the Intertemporal CAPM that allows for any state variable that predicts future investment opportunities (be aware of factor fishing!) for instance macroeconomic factors

Which factors to use?

Factor selection (cont'd)

[...]

2. Statistical: From APT we can extract factors from a large data set of asset returns, using e.g. Principal Component Analysis.
3. Firm characteristics: Creating facors based on firm characteristics, motivated by return anomalies. Most prominent example are the SMB and HML of Fama and French (1993)

- We will continue as if K factors have been chosen, for whatever of above reasons
- We will focus on unconditional asset pricing models

Fama-Macbeth

Fama-Macbeth regressions

- A pioneering approach to **estimating asset pricing models** and conducting **inference** is through Fama-Macbeth two-stage regressions.

Fama-MacBeth procedure

The Fama and MacBeth (1973) methodology is a **cross-sectional regression method** that consists of **two-steps**.

1. In the **first step**, we obtain time-series **betas** β_i for all assets from time series regressions of excess returns onto risk factors. This step is necessary as β_i is **not directly observable** and, thus, needs to be estimated.
2. In the **second step**, we obtain an estimate of the **risk premia** γ through a series of cross-sectional regressions using the estimated β_i, $\hat{\beta}_i$, from the first step as input.

Fama-MacBeth regressions (cont'd)

Fama-MacBeth procedure: First-stage regression

For each asset $i = 1, \ldots, N$, we estimate β_i using a single full sample **time-series OLS regression** of the form

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i f_t + \varepsilon_{i,t}, \tag{1}$$

where $\varepsilon_{i,t}$ is a zero-mean error term.

$\ldots$, obtaining $\hat{\beta}_i$ for $i, \ldots, N$.

- We use a **full-sample estimation** in obtaining $\hat{\beta}_i$, effectively assuming a **constant factor loading** (risk exposure). This can be extended to **rolling** βs which we will see below.

Fama-MacBeth regressions (cont'd)

Fama-MacBeth procedure: Second-stage regression(s)

For each time-period $t = 1, \ldots, T$, we run **cross-sectional regressions** of all assets against **the estimated betas**, i.e.

$$R_{i,t} - R_{f,t} = \gamma_{0t} + \gamma_t \hat{\beta}_i + \eta_{i,t}, \tag{2}$$

where the estimated value $\hat{\beta}_i$ is obtained from the **first-stage time series regressions** and η_{it} is a mean-zero error term.
Estimating these T cross-sectional regressions provides us with a **time series of estimates** of $\{\hat{\gamma}_{0,t}, \hat{\gamma}_t\}$, which can be used to form estimates of $\hat{\gamma}_0$ and $\hat{\gamma}$ as follows

$$\hat{\gamma}_j = \frac{1}{T} \sum_{t=1}^{T} \hat{\gamma}_{j,t}, \tag{3}$$

where $j = \{0, 1, \ldots, K\}$.

Fama-MacBeth regressions (cont'd)

- Now we have an **estimate** for the **risk premia** of the factors.
- Fama and MacBeth (1973) suggest the following expression for computing an **estimate of the variance** of each risk premium estimate

$$\text{Var}[\hat{\gamma}_j] = \frac{1}{T^2} \sum_{t=1}^{T} \left(\hat{\gamma}_{jt} - \hat{\gamma}_j\right)^2. \tag{4}$$

- One can also get the entire covariance matrix for risk premia (useful later in this lecture), here assuming a constant is subsumed in γ, by

$$\text{Var}[\hat{\gamma}] = \frac{1}{T^2} \sum_{t=1}^{T} \left(\hat{\gamma}_t - \hat{\gamma}\right)\left(\hat{\gamma}_t - \hat{\gamma}\right)'. \tag{5}$$

Fama-MacBeth regressions (cont'd)

Fama-MacBeth procedure: Hypothesis testing

With an estimate $\hat{\gamma}_j$ and an associated standard error $\sqrt{\mathrm{Var}[\hat{\gamma}_j]}$ in hand, our interest is in testing the following null hypothesis

$$H_0 : \gamma_j = 0. \tag{6}$$

We can test this hypothesis using the following (conventional) t-statistic

$$t(\gamma_j) = \frac{\hat{\gamma}_j}{\sqrt{\mathrm{Var}[\hat{\gamma}_j]}} \xrightarrow{d} N(0,1), \quad \text{as } T \to \infty. \tag{7}$$

The test statistic $t(\gamma_j)$ follows a student-t distribution with $T - K$ degrees of freedom in finite samples and a standard normal distribution asymptotically (why?).

…One can naturally entertain many different hypotheses using this framework.

Rolling βs as an alternative first-stage

- In fact, the original Fama-Macbeth methodology was introduced with **rolling window** βs.

Rolling βs

- One way to obtain (some) **time-variation** into β is to run a **first-stage type regression** at every time point t for $t = M, \ldots, T$, where M is the length of the rolling window.
- Typically, the **window length is fixed**, discarding the oldest time point whenever data in the most recent one gets available.
- This generates $\hat{\beta}_{it}$.

- Inclusion in Fama-Macbeth regressions is simple, as it amounts to using $\hat{\beta}_{it}$ for the t'th cross-sectional regression instead of always $\hat{\beta}_i$.

A single cross-sectional regression

- It can suffice to run a **single cross-sectional regression** in the **second stage**, which is the general approach explained in Goyal (2012) (cf. his eq. (18)).

Single cross-sectional regression

The single cross-sectional regression approach estimates

$$\overline{R_{it} - R_{ft}} = \gamma_0 + \gamma \hat{\beta}_i + \eta_i, \tag{8}$$

$\overline{R_{it} - R_{ft}} = T^{-1} \sum_{t=1}^{T} R_{it} - R_{ft}$ is the sample average of excess return to asset i.

- This is motivated from **rational expectations of investors** (or that $\mathbb{E}[R_{it} - R_{ft}]$ can be consistently estimated by its sample average, given stationarity of data).
- This will provide us with estimates of $\hat{\gamma}_0$ and $\hat{\gamma}$ that are identical to those from the t-by-t procedure.
- The usual **OLS standard errors** will be **very very** wrong.

Example: Linear consumption-based asset pricing

- One can obtain an **implementable, linear consumption-based factor model** without the need for specifying a certain utility function as per

$$\mathbb{E}[R_{it} - R_{ft}] = \gamma_c \beta_{ic}, \tag{9}$$

  where "c" indicates consumption growth, denoted $\tilde{c}_t$, and

$$\beta_{ic} = \frac{\text{Cov}[R_{it} - R_{ft}, \tilde{c}_t]}{\text{Var}[\tilde{c}_t]}. \tag{10}$$

- The ($\approx$ 2 page) derivations can be found in my lecture notes.
- Let us consider an implementation in the Matlab live script *famaMacbeth.mlx*.

Drawback of the Fama-MacBeth approach

* The Fama-MacBeth approach, while simple and highly useful, does have several problems

1. The inputs, $\mathbb{E}\left[\widetilde{r}_i\right]$, $\mathbb{E}\left[\widetilde{r}_M\right]$ and β_{iM}, are unobservable and have to be estimated
2. This gives rise to a so-called errors-in-variables problem as we are using estimated β_{iM}s in the second-stage cross-sectional regression
   - The errors-in-variables problem biases standard errors and biases λ_M towards zero
   - To reduce the problem, one can either group stocks into portfolio to get better β_{iM} estimates
   - or we can explicitly adjust standard errors to account for the bias introduced by the errors-in-variables problem (outside the scope of this course)
3. The approach does not account for autocorrelation and heteroskedasticity
   - One can solve this by using GMM instead (outside the scope of this course)

Errors-in-variables

- A major problem with the conventional Fama-Macbeth regression analysis is a **generated regressor** or **errors-in-variables** (EIV) issue.
- In the **cross-sectional stage**, **explanatory variables** are themselves **estimates** and contain, therefore, **estimation error**
- This estimation error, $v_i = \hat{\beta}_i - \beta_i$, will cause an **overstated precision** of the **risk premia estimates** if one uses the classical Fama-Macbeth standard errors from (4)

Errors-in-variables

- The **overstated precision** is directly a function of the **precision** in $\hat{\beta}$s. As such:
  1. Macroeconomic data are typically **measured with error** and often **weakly related to returns**, causing β to be imprecisely estimated
  2. If **risk factors are returns** themselves, this will generally reduce estimation error in $\hat{\beta}$
  3. The **larger time-series** the less estimation error in $\hat{\beta}$. For instance, monthly frequency tends to deliver less problems with EIV than annual data
  4. **Portfolios** of returns typically **average out noise** from individual assets, hence using those as test assets (LHS in first stage regression) improves precision of $\hat{\beta}$

Shanken correction for errors-in-variables

- Shanken (1992) provides a **solution** to the **EIV problem**
- OLS standard errors are scaled upwards to reflect this overstated precision of $\hat{\gamma}$.
- The **correction term** depends on which version of the Fama-Macbeth regression analyses is applied

Shanken corrections for EIV

Let $\text{Var}[\hat{\gamma}]$ be the Fama-Macbeth covariance matrix given in (4), thus including the intercept. Then the Shanken-corrected covariance matrices are as follows:

1. If one uses full-sample (constant) βs from the first stage in a t-by-t cross-sectional stage,

$$\text{Var}_{\text{EIV}}[\hat{\gamma}] = T^{-1}\left((1+c)\left(T\text{Var}[\hat{\gamma}] - \widetilde{\text{Var}}[f_t]\right) + \widetilde{\text{Var}}[f_t]\right) \quad (11)$$

[...]

Shanken correction for errors-in-variables

Shanken corrections for EIV (cont'd)

[...]

2. If one uses **full-sample** (constant) βs from the first stage in a **single** cross-sectional stage,

$$\text{Var}_{\text{EIV}}[\hat{\gamma}] = T^{-1}\left((1+c)T\text{Var}[\hat{\gamma}] + \widetilde{\text{Var}}[f_t]\right) \tag{12}$$

3. If one uses **rolling** βs, estimated over y years with m data points per year, from the first stage in a t-by-t cross-sectional stage,

$$\text{Var}_{\text{EIV}}[\hat{\gamma}] = T^{-1}\left((1+c^*)T\text{Var}[\hat{\gamma}] + \widetilde{\text{Var}}[f_t]\right) \tag{13}$$

where $c = \hat{\gamma}'\widetilde{\text{Var}}[f_t]^{-1}\hat{\gamma}$, $\text{Var}[f_t]$ the sample covariance matrix of the risk factors, $\widetilde{\text{Var}}[f_t]$ the $(K+1)\times(K+1)$ matrix with zeros in the first row and column, corresponding to the places of the intercept, and $\text{Var}[f_t]$ in the lower right block, and

$$c^* = c\left(1 - \frac{(y-1)(y+1)}{3yT/m}\right). \tag{14}$$

Shanken correction for errors-in-variables

- Be **careful** in using the **correct formulas** - this is not always acknowledged in the empirical literature ((13) is hidden in a footnote ... in an Appendix!)
- The effect of the Shanken correction can be substantial and impact conclusions severely, see e.g. Table II of Kan et al. (2013).

Shanken correction for errors-in-variables

Panel A: OLS										
	CAPM		C-LAB				FF3			
	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{lab}$	$\hat{\gamma}_{prem}$	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{smb}$	$\hat{\gamma}_{hml}$
Estimate	1.61	−0.46	1.77	−0.90	0.21	0.45	1.94	−0.95	0.16	0.41
t-ratio$_{fm}$	3.48	−1.19	4.16	−2.48	1.76	3.53	5.64	−3.00	1.18	3.41
t-ratio$_{s}$	3.46	−1.18	2.63	−1.70	1.12	2.25	5.45	−2.93	1.18	3.41
t-ratio$_{jw}$	3.39	−1.17	2.79	−1.79	1.20	2.46	5.53	−2.93	1.19	3.44
t-ratio$_{pm}$	3.12	−1.11	2.78	−1.76	0.99	2.71	5.17	−2.75	1.19	3.42
	ICAPM						CCAPM			
	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{term}$	$\hat{\gamma}_{def}$	$\hat{\gamma}_{div}$	$\hat{\gamma}_{rf}$	$\hat{\gamma}_0$	$\hat{\gamma}_{cg}$		
Estimate	1.14	−0.15	0.20	−0.14	−0.02	−0.44	0.96	0.18		
t-ratio$_{fm}$	2.61	−0.47	2.50	−2.69	−1.32	−3.13	2.57	0.75		
t-ratio$_{s}$	1.69	−0.33	1.62	−1.75	−0.89	−2.03	2.51	0.73		
t-ratio$_{jw}$	1.76	−0.35	1.56	−1.55	−0.91	−1.84	2.53	0.76		
t-ratio$_{pm}$	1.56	−0.32	1.38	−1.50	−0.85	−1.85	2.14	0.65		
	CC-CAY				U-CCAPM		D-CCAPM			
	$\hat{\gamma}_0$	$\hat{\gamma}_{cay}$	$\hat{\gamma}_{cg}$	$\hat{\gamma}_{cg\text{-}cay}$	$\hat{\gamma}_0$	$\hat{\gamma}_{cg36}$	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{cg}$	$\hat{\gamma}_{cgdur}$
Estimate	1.46	−1.46	−0.02	0.00	0.68	3.46	2.20	−1.18	0.45	1.84
t-ratio$_{fm}$	3.81	−2.42	−0.13	0.99	1.15	3.39	5.82	−3.48	2.26	3.25
t-ratio$_{s}$	2.62	−1.67	−0.09	0.69	0.80	2.36	4.41	−2.81	1.72	2.48
t-ratio$_{jw}$	3.26	−2.07	−0.10	0.78	0.93	2.66	5.22	−3.27	1.69	2.42
t-ratio$_{pm}$	2.86	−1.21	−0.06	0.30	0.95	2.34	5.10	−3.22	1.25	2.30

Shanken correction for errors-in-variables

- Let us consider some examples by continuing our example in the Matlab live script *famaMacbeth.mlx*.

Shanken correction for errors-in-variables

- Note that in the *famaMacbeth.mlx* example where we included **firm characteristics**, the Shanken correction **needs** a slight augmentation to function properly
- The typical argument is that **firm characteristics** are **directly observed** and **does not contribute** to **EIV problems** in the cross-sectional stage
- For that reason, we (still) only need to incorporate the EIV issues coming from estimated βs which influence the risk premia associated with characteristics
- This is achieved by augmenting $\widetilde{\text{Var}}[f_t]$ as to **include zero rows and zero columns** at the **place of the characteristics**, similarly to what we did for the the intercept

GMM

GMM approach to cross-sectional asset pricing

- The Fama-Macbeth regression analysis, with or without Shanken corrections, still **does not account for the presence of autocorrelation**
- Moreover, they tend to **assume normaility of regression errors** as well
- Lastly, while the approach by Shanken (1992) appears manageable to correct for EIV, an **easier** and **much more elegant** approach exists

$\Rightarrow$ map the whole thing into **GMM**!

GMM approach to cross-sectional asset pricing

GMM approach (intuitively)

- The main idea is to define two sets of moments:
    1. The first set matches the time series stage for obtaining β
    2. The second set matches the cross-sectional stage for obtaining γ
- Merging those two sets of moments **"internalizes"** any EIV, as β and γ are essentially estimated jointly
- The long-run covariance matrix of moments S will then **capture** directly the **effect of generated regressors**
- …and we already know how to amend S (nonparametrically) to account for autocorrelation and heteroskedasticity through HAC

GMM approach to cross-sectional asset pricing

Definition: Beauty

The beauty of this approach is that it captures (almost) all issues in one set of moments, yet it is essentially **still a Fama-Macbeth regression** analysis but with an additional layer that provides proper and accurate standard errors accounting for both EIV, autocorrelation, and heteroskedasticity and is much more mild in assumptions.

GMM approach to cross-sectional asset pricing

- For simplicity, let us define excess returns for the i'th asset as $R_{it}^e \equiv R_{it} - R_{ft}$, $i = 1, \ldots, N$.
- We may write the time series stage in vector form as follows

$$R_t^e = \alpha + \beta f_t + \varepsilon_t, \tag{15}$$

  where $R_t, \alpha, \varepsilon_t$ are all $N \times 1$, β is $N \times K$ and f_t is $K \times 1$.
- Now, recall that the **identifying moment conditions** of OLS are, intuitively, that the error term is mean zero and that it is uncorrelated with the regressors.

GMM approach to cross-sectional asset pricing

- Formally, the **OLS moment conditions** for a univariate dependent variable y, regressors x, and coefficients θ are

$$\mathbb{E}[y - \theta x] = 0 \quad \text{and} \quad \mathbb{E}[(y - \theta x)x] = 0. \tag{16}$$

Time-series stage moment conditions

Thus, OLS estimation of the time-series stage maps into the following set of moments

$$\mathbb{E}\begin{bmatrix} R_t^e - \alpha - \beta f_t \\ (R_t^e - \alpha - \beta f_t) \otimes f_t \end{bmatrix} = \mathbb{E}\begin{bmatrix} \varepsilon_t \\ \varepsilon_t \otimes f_t \end{bmatrix} = \begin{bmatrix} 0_{N \times 1} \\ 0_{NK \times 1} \end{bmatrix}, \tag{17}$$

equalling $N + NK$ moment conditions.

- Since we have N many αs and NK many βs to estimate, the system is exactly identified and **reduces to the analytical solution of OLS regressions** for each i
- The Kronecker product ensures that all errors terms are uncorrelated with all risk factors

GMM approach to cross-sectional asset pricing

- Recall that the β representation of the K-factor asset pricing model delineates

$$\mathbb{E}[R_t^e] = \gamma_0 + \gamma\beta. \tag{18}$$

- Strictly speaking, in many cases there is **no constant** γ_0 and the restriction $\gamma_0 = 0$ could be imposed
- We will focus on the case where we include the constant and consider $\gamma_0 = 0$ a **testable restriction**
- This defines the second, **cross-sectional stage of the Fama-Macbeth analysis**
- As such, it will naturally also define the second set of moments…

GMM approach to cross-sectional asset pricing

Cross-sectional stage moment conditions

Thus, the implication from any K-factor asset pricing model expressed in β representation is the following moment conditions

$$\mathbb{E}[R_t^e - \gamma_0 - \gamma\beta] = 0_{N\times 1}. \tag{19}$$

Joint moment conditions

Thus, the joint moment conditions are given by

$$\mathbb{E}\begin{bmatrix} R_t^e - \alpha - \beta f_t \\ (R_t^e - \alpha - \beta f_t) \otimes f_t \\ R_t^e - \gamma_0 - \gamma\beta \end{bmatrix} = \mathbb{E}\begin{bmatrix} \varepsilon_t \\ \varepsilon_t \otimes f_t \\ \eta_i \end{bmatrix} = \begin{bmatrix} 0_{N\times 1} \\ 0_{NK\times 1} \\ 0_{N\times 1} \end{bmatrix}, \tag{20}$$

equalling $N(K+2) = NK + 2N$ moment conditions.

- Note that the system is now **overidentified**, as we added N moments and only need to estimate $K+1$ additional parameters.

GMM approach to cross-sectional asset pricing

- The joint system in (20) will generally **not** reproduce OLS estimates. In order to achieve those (and be consistent with the Fama-Macbeth approach), we need to amend the moments slightly.
- Similarly to the time-series stage, we need mean zero errors and them being uncorrelated with regressors.
- To achieve that, define the following $(N+1)(K+1) \times N(K+2)$ matrix

$$e = \begin{bmatrix} I_{N(K+1)} & 0_{N(K+1)\times N} \\ 0_{(K+1)\times N(K+1)} & \chi' \end{bmatrix}, \tag{21}$$

  where $\chi = (\iota_{N\times 1}, \beta)$ is $N \times (K+1)$ and ι is a N-vector of ones.
- Note that $(N+1)(K+1) = NK + N + K + 1$.

GMM approach to cross-sectional asset pricing

- The neat thing here is that when **pre-multiplying** e onto the joint moment conditions in (20), we maintain the first $N + NK$ conditions as is (the time series stage moments), but weight each of the last N moments conditions (the cross-sectional stage moments) by χ
- $\ldots \chi$ is indeed containing the regressors used in the cross-sectional stage
- …so it fits the natural OLS type of identifying moments!

Joint OLS-type moment conditions

The OLS type moment conditions

$$\begin{bmatrix} I_{N(K+1)} & 0_{N(K+1)\times N} \\ 0_{(K+1)\times N(K+1)} & \chi' \end{bmatrix} \mathbb{E} \begin{bmatrix} R_t^e - \alpha - \beta f_t \\ (R_t^e - \alpha - \beta f_t) \otimes f_t \\ R_t^e - \gamma_0 - \gamma\beta \end{bmatrix} = \begin{bmatrix} 0_{N\times 1} \\ 0_{NK\times 1} \\ 0_{(K+1)\times 1} \end{bmatrix},$$

which is equivalent to $eg = 0_{(N+1)(K+1)}$ and where g is defined as (20), reproduces the two-pass Fama-Macbeth estimates.

GMM approach to cross-sectional asset pricing

- So **why bother** doing all this if we might as well just run Fama-Macbeth regressions? **standard errors**!
- Of course, we could also estimate parameters directly via the standard GMM recipe, yet why not keep it simple?

The GMM recipe for asset pricing

In order to **estimate and evaluate asset pricing models** in a β representation, the following approach will be used:

1. Estimate β_i for $i = 1, \ldots, N$ via standard Fama-Macbeth time series stage regressions.
2. Estimate γ_j for $j \in \{0, 1, \ldots, K\}$ via a standard Fama-Macbeth single cross-sectional regression.
3. Use the definition of e in (21) and the joint moment condition system in (20) to obtain standard errors that are robust to EIV, autocorrelation, and heteroskedasticity.

GMM approach to cross-sectional asset pricing

- To compute those standard errors, define

$$\theta' = (\alpha', \mathrm{vec}(\beta)', \gamma_0, \gamma)', \tag{22}$$

  which contains all $N + NK + 1 + K = (N+1)(K+1)$ parameters to be estimated.
- The vectorization defines the transformation of any matrix of dimension, say, $N \times K$, into a column vector of dimension $NK \times 1$ simply by stacking all columns of the matrix on top of another.
- We are interested in obtaining $\mathrm{Var}[\hat{\theta}]$ for which we need the ingredients of e, S, and yet another matrix (the gradient) defined on the next slide.

GMM approach to cross-sectional asset pricing

Covariance of $\hat{\theta}$

The $(N+1)(K+1) \times (N+1)(K+1)$ covariance of $\hat{\theta}$ is

$$\text{Var}[\hat{\theta}] = T^{-1}(eD)^{-1} eSe'(eD)^{-1\prime}, \tag{23}$$

where $D = \mathbb{E}[\partial g(\theta)/\partial\theta]$ is equal to

$$D = -\begin{bmatrix} \begin{bmatrix} 1 & \mathbb{E}[f_t'] \\ \mathbb{E}[f_t] & \mathbb{E}[f_t f_t'] \end{bmatrix} \otimes I_N & 0_{N(K+1)\times(K+1)} \\ \begin{bmatrix} 0 & \gamma' \end{bmatrix} \otimes I_N & \chi \end{bmatrix} \tag{24}$$

and S is the long-run covariance matrix equal to

$$S = \sum_{s=-\infty}^{\infty} \mathbb{E}\left[\begin{bmatrix} R_t^e - \alpha - \beta f_t \\ (R_t^e - \alpha - \beta f_t) \otimes f_t \\ R_t^e - \gamma_0 - \gamma\beta \end{bmatrix} \begin{bmatrix} R_{t-s}^e - \alpha - \beta f_{t-s} \\ (R_{t-s}^e - \alpha - \beta f_{t-s}) \otimes f_{t-s} \\ R_{t-s}^e - \gamma_0 - \gamma\beta \end{bmatrix}' \right] \tag{25}$$

(Note that there is an error in D in Goyal (2012))

GMM approach to cross-sectional asset pricing

- Note that the formula is just the formula shown in past lectures with a certain weighting matrix e.
- …and S is defined similarly as earlier, so is D.
- It is neat that D has an analytical form, and $\mathbb{E}[f_t]$ and $\mathbb{E}[f_t f_t']$ can easily be estimated by their sample counter parts as

$$T^{-1} \sum_{t=1}^{T} f_t \xrightarrow{p} \mathbb{E}[f_t] \tag{26}$$

and

$$T^{-1} \sum_{t=1}^{T} f_t f_t' \xrightarrow{p} \mathbb{E}[f_t f_t']. \tag{27}$$

GMM approach to cross-sectional asset pricing

- The last ingredient is an estimate of S.
- We discussed this during the second double lectures, including how to construct a nonparametric HAC we did this for any generic set of moment conditions, in which (20) naturally falls.
- Let us consider how all this can be implemented, using the Matlab live script *GMM_crossSectionalAssetPricing.mlx*.

A comment on traded vs. non-traded factors

- Some factors are returns themselves, e.g. the market risk premium, SMB, or HML.
- In those cases, we do not need to estimate their risk premium in a cross-sectional stage, but can suffice by taking its sample mean.
- When factors are non-traded, we need to estimate their risk premia using the cross-sectional stage.

Why would one do the cross-sectional stage with traded factors then?

Lewellen et al. (2010) emphasize that one useful diagnostic test of an asset pricing model with traded factors (e.g. CAPM or the Fama-French three factor model) is that the estimated risk premia from the cross-sectional stage should be statistically indistinguishable from their sample mean taken over the time series dimension.

Estimating SDF loadings

- Instead of estimating the β representation, we could also have estimated the parameters of the SDF (called SDF loadings) denoted by b.
- We can make inference on them, and back out risk premia directly through the covariance matrix of factors.
- But be aware that the statement by Goyal (2012) that

  "Which method one uses [TSR+CSR approach or SDF approach] is, therefore, largely a matter of individual preference."

  is at best very imprecise and slightly wrong.
- Even though risk premia and SDF loading are directly related, they differ in interpretation!

Estimating SDF loadings

- For now, we will state the approach, assuming an SDF of the form

$$M_t = 1 - b' f_t. \tag{28}$$

- We may then form standard GMM moment conditions from the fundamental equation of asset pricing as

$$\mathbb{E}[(1 - b' f_t) R_t^e] = 0_{N \times 1}. \tag{29}$$

- We also have that (be careful with the dimensions in Goyal (2012) here)

$$D = \mathbb{E}\left[R_t^e f_t'\right], \tag{30}$$

which can be estimated simply as

$$T^{-1} \sum_{t=1}^{T} R_t^e f_t'. \tag{31}$$

Estimating SDF loadings

- Note also that

$$S = \sum_{s=-\infty}^{\infty} \mathbb{E}[(R_t^e - b' f_t R_t^e)(R_{t-s}^e - b' f_{t-s} R_{t-s}^e)'], \tag{32}$$

where Goyal (2012) has a minor typo as well in his eq. (40).

SDF loading estimates and their variance

If the weighting matrix is the identity matrix (yielding $\hat{b}_1$) or the optimal S^{-1} matrix (yielding $\hat{b}_2$), we find

$$\hat{b}_1 = (D'D)^{-1} D' \overline{R^e}, \tag{33}$$

$$\hat{b}_2 = (D'S^{-1}D)^{-1} D' S^{-1} \overline{R^e}. \tag{34}$$

with

$$\text{Var}[\hat{b}_1] = T^{-1}(D'D)^{-1} D'SD(D'D)^{-1}, \tag{35}$$

$$\text{Var}[\hat{b}_2] = T^{-1}(D'SD)^{-1}. \tag{36}$$

Skeptical appraisal

A skeptical appraisal of asset pricing tests

- While our approach to **cross-sectional asset pricing** is **intriguing** and deals with numerous econometric issues, it is still subject to critique.

A skeptical appraisal

In many situations encountered in practice, it may be easy to find factors that explain the cross-section of expected returns. **Finding a high cross-sectional R^2 and small pricing errors often has little economic meaning** and, in the authors' view, does not, by itself, provide much support for a proposed model. The problem is not just a sampling issue - it cannot be solved by getting standard errors right - though sampling issues exacerbate the problem.

- Lewellen et al. (2010) delineate several **"prescriptions"** as to how to conduct proper asset pricing analyses.
- Note, however, the paper is from 2010 and many improvements have since then been developed several of which we will see in Weeks 8 and 9.

A skeptical appraisal of asset pricing tests

Prescription 1

Expand the set of test portfolios beyond size-value portfolios of Fama and French (1993), for instance using industry portfolios, statistical portfolios, additional asset classes (bonds, currencies, etc.) or simply use individual stocks.

- In many, many applications the only set of assets used in testing a given asset pricing model or whether a given factor is priced is the 25 size-value portfolios of Fama and French (1993)
- However, their cross-sectional variation exhibit a very strong factor structure by **by construction**, well explained by a few factors (SMB and HML)
- It is, as such, not very challenging to find any other factor that explains the cross-sectional variation in those assets.
- …a related solution is to add SMB and HML (or ME and BM characteristics) to your model and test whether they drive out your new factor(s)/model

A skeptical appraisal of asset pricing tests

Prescription 2

Take the magnitude, sign, and significance of the cross-sectional coefficients seriously.

- If a model implies $\gamma_0 = 0$ (like the CAPM), **make sure that you test** this and comment on the results, even though it provides a high R^2.
- If a model implies that γ is equal to **average factor excess returns** (for traded factors), make sure you evaluate this. Otherwise, it is **indication of model misspecification**

A skeptical appraisal of asset pricing tests

Prescription 3

Report confidence intervals for the cross-sectional R^2.

- While we will not deal with asymptotic distributions of R^2 in this course, the prescription is still important for the interpretation of R^2.
- …that is, R^2 is also an **estimated** metric and should be considered as such
- It typically has very wide confidence bands, representing this kind of uncertainty on the pricing ability, see e.g. Kan et al. (2013)

Other good habits

- Make a **sufficient amount of robustness checks** using, e.g., other sample periods or test assets and generally challenge the subjective choices you have made in the analysis.

A (repeated) message

…most importantly, **always** have a strong economic motivation for why the models works/makes sense.

Potential projects

About potential projects

- Cross-sectional asset pricing deals with questions like:
  1. What explains the cross-sectional variation in expected returns?
  2. Which risk factors matter? What are their (required) compensation in the financial market?
  3. Is a certain risk factor priced in the financial markets?
- Of course, all analyses have a certain focus, for instance the paper by Menkhoff et al. (2012) that posed the question as to whether FX volatility risk explained the cross-sectional variation in average return on carry trades.
- Or Fama and French (1993) that addressed the puzzle that CAPM fails by proposing two new factors.

About potential projects

- Test the cross-sectional asset pricing abilities of a certain risk factor or model, preferably a *new* risk factor.
- Test an existing model on *new* asset classes, like currencies, bonds, cryptocurrencies, commodities etc.
- Re-evaluate an existing and important, yet likely outdated model or risk factor proposed in the literature with new, up-to-date data, more subsamples, etc.
- Test a conditional model via scaled factors, like Lettau and Ludvigson (2001). For instance, recession attention from Bybee et al. (2019), EPU Baker et al. (2016), climate policy uncertainty Gavriilidis (2021), etc...
- …many of these ideas can be merged with those coming from the next topic

References

BAKER, S. R., N. BLOOM, AND S. J. DAVIS (2016): "Measuring economic policy uncertainty," *The quarterly journal of economics*, 131, 1593–1636.

BORUP, D. AND E. C. M. SCHÜTTE (2021): "Asset pricing with data revisions," *Journal of Financial Markets*, Forthcoming.

BYBEE, L., B. T. KELLY, A. MANELA, AND D. XIU (2019): "The structure of economic news," Working paper.

DELIKOURAS, S. AND A. KOSTAKIS (2019): "A single-factor consumption-based asset pricing model," *Journal of Financial and Quantitative Analysis*, 54, 789–827.

FAMA, E. F. AND K. R. FRENCH (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

FAMA, E. F. AND J. D. MACBETH (1973): "Risk, return, and equilibrium: Empirical tests," *Journal of Political Economy*, 81, 607–636.

GAVRIILIDIS, K. (2021): "Measuring climate policy uncertainty," *Working Paper*.

GOYAL, A. (2012): "Empirical cross-sectional asset pricing: a survey," *Financial Markets and Portfolio Management*, 26, 3–38.

KAN, R., C. ROBOTTI, AND J. SHANKEN (2013): "Pricing model performance and the two-pass cross-sectional regression methodology," *The Journal of Finance*, 68, 2617–2649.

LETTAU, M. AND S. LUDVIGSON (2001): "Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying," *Journal of political economy*, 109, 1238–1287.

LEWELLEN, J., S. NAGEL, AND J. SHANKEN (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial economics*, 96, 175–194.

MENKHOFF, L., L. SARNO, M. SCHMELING, AND A. SCHRIMPF (2012): "Carry trades and global foreign exchange volatility," *The Journal of Finance*, 67, 681–718.

SHANKEN, J. (1992): "On the estimation of beta-pricing models," *Review of Financial Studies*, 5, 1–33.

Return Predictability

Empirical Asset Pricing

Mads Markvart Kjær

Department of Economics and Business Economics, Aarhus University

E-mail: mads.markvart@econ.au.dk

Spring 2023

Do you remember?

- Expected excess return of any asset can be written as:

$$\mathbb{E}_t(R_{i,t+1}) - R_{f,t} = -\frac{\text{Cov}_t(M_{t+1}, R_{i,t+1})}{\mathbb{E}_t(M_{t+1})} \tag{1}$$

- Any variable that affects either;
  - The conditional covariance between the SDF and the return of asset i,
  - The expectation to the SDF itself

  $\rightarrow$ will predict future expected returns!
- The question is, of course, whether such variable exists? And whether (excess) returns, thereby, are predictable?

  $\rightarrow$ In this lecture, we look into the ongoing and long-standing debate about this question!

Return predictability in a nutshell

- This seemingly simple question has generated an astonishing amount of mixed empirical evidence

  A series of papers document predictability is many diverse asset markets

  - **Stocks:** Campbell and Shiller (1988), Fama and French (1989), Cochrane (2008), Campbell and Thompson (2008), Atanasov et al. (2020), Gu et al. (2020)
  - **Bonds:** Fama and Bliss (1987), Campbell and Shiller (1991), Cochrane and Piazzesi (2005), Ludvigson and Ng (2009), Cieslak and Povala (2015), Eriksen (2017), Bianchi et al. (2021)
  - **Currencies:** Molodtsova and Papel (2009), Della Corte et al. (2009), Li et al. (2015), Cheung et al. (2019)

  Another argue that predictability is spurious and performs poorly out-of-sample

  - **Stocks:** Stambaugh (1986, 1999), Nelson and Kim (1993), Goyal and Welch (2008), Goyal et al. (2021)
  - **Bonds:** Thornton and Valente (2012), Sarno et al. (2016), Bauer and Hamilton (2018), Ghysels et al. (2018)
  - **Currencies:** Meese and Rogoff (1983a), Meese and Rogoff (1983b), Rossi (2013)

Return predictability in a nutshell

Return predictability

Return predictability is typically studied using a predictive regression model

$$r_{t+1} = \alpha + \boldsymbol{x}_t' \boldsymbol{\beta} + \varepsilon_{t+1} \tag{2}$$

where r_{t+1} is the one–period ahead log excess return on the asset, x_t is a predictor variable, ε_{t+1} is a zero-mean disturbance term, and a $\beta \neq 0$ implies that returns are predictable

- If excess returns (risk premia) are predictable, as pointed out in Fama and French (1989), then expected excess returns vary over time as a function of $\boldsymbol{x}_t$

$$\mathbb{E}_t \left[r_{t+1} \right] = \widehat{\alpha} + \boldsymbol{x}_t' \widehat{\boldsymbol{\beta}} \tag{3}$$

- At its core, this approach is universally applicable across stocks, bonds, currencies, commodities, …
- For a recent and comprehensive survey of stock return predictability and how to evaluate it, see Rapach and Zhou (2013)

Why study return predictability?

- Return predictability has a long tradition in empirical asset pricing and is an all around fascinating endeavor
  1. Can help help answer a long-standing debate: Does expected excess returns (risk premia) vary over time?
  2. Can help sharpen the distinction between risk-based (market efficiency) and behavioral (ineffiency) explanations for variations in returns
  3. Results can lead to better and more realistic asset pricing models (remember the UMVE implied SDF)
  4. Results can lead to better investment performance for households, mutual funds, pension companies, and policy makers
- It can also, however, be a frustrating exercise with many issues
  1. Thorny econometric issues complicate inference
  2. (Excess) returns inherently contains a large unpredictable component
  3. How to deal with model uncertainty and instability?
  4. How to use the abundance of data available without overfitting the model?
  5. We do not know "*The Model*" or the data generating process (DGP) for returns

Which variables are we looking for?

- In general, the literature has two potential explanations:

  1. Rational risk-based (remember the relation to SDF from ealier)
  2. Frictions (and behavioal biases) that leads to market-inefficiencies

- A potential candidate can natually also reflect both explanations

Examples

Stock Prices and Wall Street Weather

By EDWARD M. SAUNDERS, JR.*

Examples

THE JOURNAL OF FINANCE • VOL. LXII, NO. 4 • AUGUST 2007

Sports Sentiment and Stock Returns

ALEX EDMANS, DIEGO GARCÍA, and ØYVIND NORLI*

Examples

JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 45, No. 2, Apr. 2010, pp. 535–553

doi:10.1017/S0022109010000153

Exploitable Predictable Irrationality: The FIFA World Cup Effect on the U.S. Stock Market

Guy Kaplanski and Haim Levy*

Examples

Journal of Financial Economics 145 (2022) 234–254

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Music sentiment and stock returns around the world ☆

Alex Edmans [a,*], Adrian Fernandez-Perez [b], Alexandre Garel [c], Ivan Indriawan [b]

[a] *London Business School, Regent's Park, London NW1 4SA, United Kingdom*
[b] *Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand*
[c] *Audencia Business School, 8 Route de la Jonelière, Nantes 44312, France*

Examples

MANAGEMENT SCIENCE
Vol. 60, No. 7, July 2014, pp. 1772–1791
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2013.1838

Forecasting the Equity Risk Premium: The Role of Technical Indicators

Christopher J. Neely
Research Division, Federal Reserve Bank of St. Louis, St. Louis, Missouri 63166, neely@stls.frb.org

David E. Rapach
Department of Economics, John Cook School of Business, Saint Louis University, St. Louis, Missouri 63108, rapachde@slu.edu

Jun Tu
Department of Finance, Lee Kong Chian School of Business, Singapore Management University, Singapore 178899, tujun@smu.edu.sg

Guofu Zhou
Olin Business School, Washington University in St. Louis, St. Louis, Missouri 63130, CAFR and CUFE, zhou@wustl.edu

Examples

The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXII, NO. 6 • DECEMBER 2017

Why Does Return Predictability Concentrate in Bad Times?

JULIEN CUJEAN and MICHAEL HASLER*

Examples

Journal of Financial Economics 134 (2019) 192–213

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

A tug of war: Overnight versus intraday expected returns☆

Dong Lou[a,b], Christopher Polk[a,b,*], Spyros Skouras[c]

[a] *Department of Finance, London School of Economics, London WC2A 2AE, UK*
[b] *CEPR, UK*
[c] *Department of International and European Economic Studies, Athens University of Economics and Business, 76 Patision St., Athens, Greece*

Frictions and behavioral biases

- The papers shown ealier relate to different explanations:
  1. Mood/sentiment of the investors
  2. Differences in perceiving information
  3. Frictions due to constraints (attention, regulation, margin requirements, etc.)
- Explanations not necerarily consistent with EMH
- We will not spend much time on these this field of literature, but now you know it exists!

A rational risk-based explanation

- The second stand is that predictability has a risk-based explanation
- This field of literature focuses on SDF from previous:

$$\mathbb{E}_t(R_{i,t+1}) - R_{f,t} = -\frac{\mathsf{Cov}_t(M_{t+1}, R_{i,t+1})}{\mathbb{E}_t(M_{t+1})} \tag{4}$$

- Meaning idenfying variables related either the expected value of the SDF or its covariance with returns
- So in most of the cases, variables are motivated from a consumption-based framework which rationalizes that the predictor of interest should predict returns

Examples

The Journal of FINANCE

The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXV, NO. 3 • JUNE 2020

Consumption Fluctuations and Expected Returns

VICTORIA ATANASOV, STIG V. MØLLER, and RICHARD PRIESTLEY*

Examples

Journal of Banking and Finance 129 (2021) 106159

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

ELSEVIER

Check for updates

Unemployment and aggregate stock returns☆

Victoria Atanasov

University of Mannheim, Chair of Finance, L9 1-2, Mannheim 68161, Germany

Examples

Journal of Financial Economics 121 (2016) 46–65

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/finec

Short interest and aggregate stock returns☆

David E. Rapach[a,1], Matthew C. Ringgenberg[b,2], Guofu Zhou[b,c,d,*]

[a] John Cook School of Business, Saint Louis University, 3674 Lindell Boulevard, St. Louis, MO 63108, USA
[b] Olin Business School, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130, USA
[c] China Academy of Financial Research, Shanghai 200030, China
[d] China Economics and Management Academy, Beijing 100081, China

Examples

Expected Stock Returns and Variance Risk Premia

Tim Bollerslev
Duke University

George Tauchen
Duke University

Hao Zhou
Federal Reserve Board

Outcome of this lecture

After the lecture, you should have

- **knowlegde** and **understading** of
  - Return predictability in financial markets, its role in asset pricing, and implications for the dynamics of asset prices
- and be able to
  - Discuss and implement in-sample and out-of-sample return predictability methods, evaluate the empirical findings, and reflect on their implications for the dynamics of asset prices

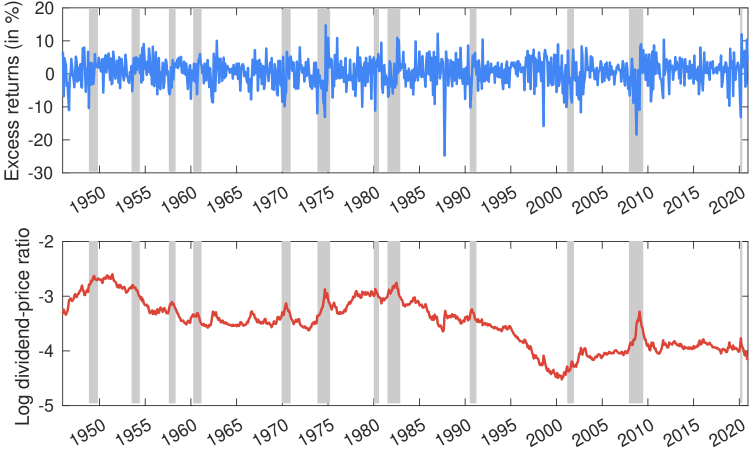
- For now, we will focus on a specific variable related to risk-based category, i.e., the dividend-price ratio
- Using this predictor, we will go the entire methodology for you to conduct an asset pricing study about time-series predictability in a consistent way

  ➜ We will see how to assess whether expected returns are constant over time or not!

In-sample predictability

S&P500 excess returns and dividend-price ratio

- We will consider the most classical example of a stock return predictor, i.e., the dividend-price ratio (Campbell and Shiller, 1988, Fama and French, 1988)



Campbell-Shiller loglinear approximation

- Campbell and Shiller (1988) propose an approximate loglinear present value model. Start from the definition of log returns

$$\begin{aligned} r_{t+1} &= \ln\left(P_{t+1} + D_{t+1}\right) - \ln\left(P_t\right) && (5) \\ &= p_{t+1} - p_t + \ln\left(1 + \exp\left(d_{t+1} - p_{t+1}\right)\right) && (6) \end{aligned}$$

- The last term is a nonlinear function of the log dividend-price ratio. Taking a first-order Taylor approximation yields the following approximation

$$r_{t+1} \approx \kappa + \phi p_{t+1} + \left(1 - \phi\right) d_{t+1} - p_t \tag{7}$$

with

$$\phi = \frac{1}{1 + \exp\left(\overline{d - p}\right)} \tag{8}$$

and

$$\kappa = -\ln\phi - \left(1 - \phi\right)\ln\left(1/\phi - 1\right) \tag{9}$$

Why the dividend-price ratio is a natural predictor

- Solving (7) forward, isolating for p_t, imposing the transversality condition $\lim_{j\to\infty} \phi^j p_{t+j} = 0$, and taking conditional expectations gives us

$$p_t = \frac{\kappa}{1-\phi} + \mathbb{E}_t \left[\sum_{j=0}^{\infty} \phi^j \left[(1-\phi)\, d_{t+1+j} - r_{t+1+j} \right] \right] \tag{10}$$

- Similarly, (7) can be stated in terms of the log dividend-price ratio and iterating forward as above yields

$$d_t - p_t = \frac{-\kappa}{1-\phi} + \mathbb{E}_t \left[\sum_{j=0}^{\infty} \phi^j \left[-\Delta d_{t+1+j} + r_{t+1+j} \right] \right] \tag{11}$$

  which tells us that *if* the dividend-price ratio varies over time, this must reflect predictable changes in either future dividends, future returns, or some combination of the two

Econometric issue: Stambaugh's finite sample bias

Stambaugh's finite sample bias

Consider the following system in which the predictor x_t follows an AR(1) process (Nelson and Kim, 1993, Kothari and Shanken, 1997, Stambaugh, 1999)

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim iid\left(0,\sigma_\varepsilon^2\right) \tag{12}$$

$$x_{t+1} = \lambda + \rho x_t + \nu_{t+1}, \quad \nu_{t+1} \sim iid\left(0,\sigma_\nu^2\right) \tag{13}$$

where ε_{t+1} and ν_{t+1} are white noise errors that are contemporaneously correlated with covariance $\sigma_{\varepsilon\nu}$

- x_t is predetermined not exogenous...
- The OLS estimate $\widehat{\rho}$ in (13) is downward-biased in a finite sample (Kendall, 1954)

$$\mathbb{E}\left[\widehat{\rho} - \rho\right] \approx -\left(\frac{1+3\rho}{T}\right) \tag{14}$$

Interpreting the bias

- Stambaugh (1986, 1999) shows that the bias in $\widehat{\beta}$ in (12) is then given by

$$\mathbb{E}\left[\widehat{\beta}-\beta\right]=\underbrace{\frac{\sigma_{\varepsilon\nu}}{\sigma_{\nu}^{2}}}_{\gamma}\mathbb{E}\left[\widehat{\rho}-\rho\right] \tag{15}$$

- Note that we can interpret the term γ in (15) as a regression coefficient

$$\varepsilon_{t+1}=\gamma\nu_{t+1}+\eta_{t+1} \tag{16}$$

- Given (15) and (14), the bias in $\widehat{\beta}$ can be approximated as

$$\mathbb{E}\left[\widehat{\beta}-\beta\right]\approx-\frac{\sigma_{\varepsilon\nu}}{\sigma_{\nu}^{2}}\left(\frac{1+3\rho}{T}\right)=-\gamma\left(\frac{1+3\rho}{T}\right) \tag{17}$$

- If ε_{t+1} and ν_{t+1} are negatively correlated, then $\gamma<0$ and the downward bias in $\widehat{\rho}$ produces an upward bias in $\widehat{\beta}$. The bias in increasing in γ and ρ, but decreasing in the sample size T
- This is the case for the dividend-price ratio (d_t-p_t), i.e., an unexpected increase in p_{t+1} leads to a negative ν_{t+1} and an unexpected increase in r_{t+1} and therefore a positive ε_{t+1}

Parametric bootstrap for valid inference

- Despite the small sample bias, we can still conduct valid inference using, e.g., bootstrapping techniques
- The idea in a nutshell: Evaluate the biased coefficient in an empirical (finite sample) distribution which suffers from the same bias $\Rightarrow$ valid inference!
- Suppose, as above, that the vector of errors $(\varepsilon_{t+1}, \nu_{t+1})'$ is multivariately normally distributed with covariance matrix

$$\begin{bmatrix} \varepsilon_{t+1} \\ \nu_{t+1} \end{bmatrix} \sim iid \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma \right), \quad \text{where} \quad \Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\nu} \\ \sigma_{\nu\varepsilon} & \sigma_\nu^2 \end{bmatrix} \tag{18}$$

- We also assume that $\rho < 1$ to ensure covariance stationarity of the regressor, but we do allow it to be highly persistent (i.e., ρ close to unity)

Parametric bootstrap example

- Suppose that we wish to test the null $\mathcal{H}_0 : \beta = \beta_0$. A parametric bootstrap would entail the following steps

1. Use OLS to obtain estimates of $(\alpha, \beta, \lambda, \rho, \Sigma) \rightarrow \left(\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda}, \widehat{\rho}, \widehat{\Sigma}\right)$
2. Generate T random numbers of $(\varepsilon_{t+1}, \nu_{t+1})$ from a multivariate normal distribution with covariance matrix $\widehat{\Sigma} \rightarrow \left(\varepsilon^*_{t+1}, \nu^*_{t+1}\right)$
3. Generate a random initial value of x_t as

$$x_1^* \sim \mathcal{N}\left(\overline{x}, \widehat{\sigma}_x^2\right) \tag{19}$$

where $\overline{x}$ and $\widehat{\sigma}_x^2$ denote the unconditional mean and variance of the predictor variable x_t, respectively

Parametric bootstrap example

4. Use $\widehat{\lambda}$ and $\widehat{\rho}$ together with the generated values of ν_{t+1} and the initial value in steps 2 and 3 to obtain T observations of x_t

$$\widehat{\lambda} + \widehat{\rho} x_t^* + \nu_{t+1}^* \rightarrow x_{t+1}^* \tag{20}$$

5. Use $\widehat{\alpha}$ and the hypothesized value β_0 together with the generated values of ε_{t+1} and x_t to obtain T observations of r_{t+1}

$$\widehat{\alpha} + \beta_0 x_t^* + \varepsilon_{t+1}^* \rightarrow r_{t+1}^* \tag{21}$$

6. Estimate the predictive regression in (12) on the simulated data to obtain $\widetilde{\beta}^{(1)}$

7. Repeat steps 2–6 M times to obtain $\widetilde{\beta}^{(1)}, \widetilde{\beta}^{(2)}, \ldots, \widetilde{\beta}^{(M)}$

- From the empirical distribution of $\widetilde{\beta}^{(i)}$, we can then compute the (upper) ones-sided p–value under the null hypothesis as

$$\mathbb{P}\left[\widetilde{\beta} > \widehat{\beta}\right] = \frac{1}{M} \sum_{i=1}^{M} \mathbf{1}\left\{\widetilde{\beta}^{(i)} > \widehat{\beta}\right\} \tag{22}$$

Parametric bootstrap example

- Alternatively, one can compute the correct critical value from the relevant percentiles of the distribution of $\widetilde{\beta}^{(i)}$

Bias in the bootstrap approach

- The bootstrap approach automatically deals with bias since, relative to the distribution under the null hypothesis, the mean of the empirical distribution moves (either to the left or the right depending on the sign of the bias) with the size of the bias

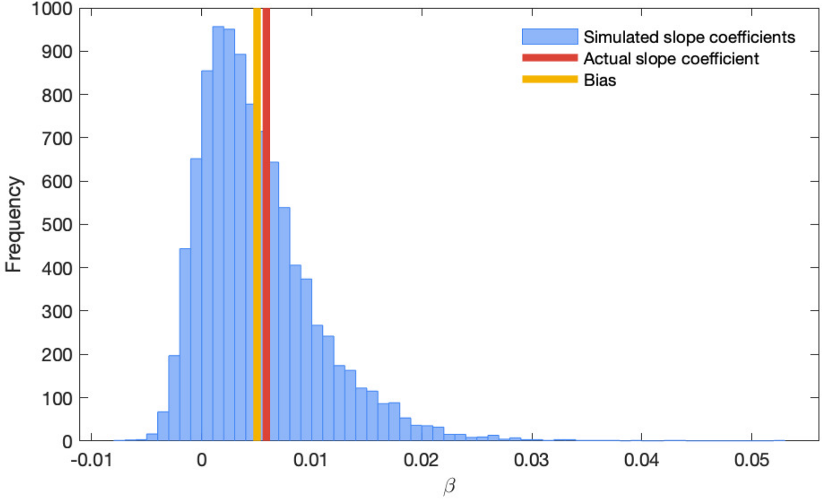
$$\text{Bias}\left(\widehat{\beta}\right) = \frac{1}{M}\sum_{i=1}^{M}\widetilde{\beta}^{(i)} - \beta_0 \tag{23}$$

- A residual-based bootstrap does not rely on a distributional assumption in generating random numbers, but instead samples from the actual data, i.e., we follow the same scheme as outlined above but replace steps 2 and 3 with

  2.* Generate T values of $(\varepsilon_{t+1}, \nu_{t+1})$ by random draws (with replacement) from the set of residuals obtained in step 1

  3.* Generate a random initial value of x_t by a random draw (with replacement) from the observed predictor variable

Bootstrapping in-sample regression



In-sample regressions and inference

- Below we consider empirical results for a study using the log dividend-price ratio to predict monthly excess stock returns
- We present full sample estimates of the system in (12) and (13) along with the bias and bootstrapped p-value for the slope parameter and Lewellen (2004) estimates (see next slide)

	Excess returns			Dividend-price ratio		
	α	β	R^2	λ	ρ	R^2
Estimate	0.0261	0.0059	0.37	-0.0136	0.9964	99.04
	(2.36)	(1.87)		(-1.20)	(312.34)	
Bias in slope		0.0050				
Bootstrap p-value		[0.36]				
Lewellen β		0.0024				
Lewellen t-stat		[4.00]				

Lewellen's (2004) defence

- Lewellen (2004) begins by conditioning the Stambaugh bias on the estimated persistence $\widehat{\rho}$ and the true persistence ρ

$$\mathbb{E}\left[\widehat{\beta} - \beta | \widehat{\rho}, \rho\right] = \gamma \left[\widehat{\rho} - \rho\right] \tag{24}$$

- At first, this may not seen particularly useful as ρ is unknown. However, since $d_t - p_t$ is not explosive, $\rho = 1$ is where the maximum bias occurs, giving us

$$\widehat{\beta}_{\text{Adj.}} = \widehat{\beta} - \gamma \left[\widehat{\rho} - 1\right] \tag{25}$$

with variance equal to (regardless of the true of ρ)

$$\text{Var}\left[\widehat{\beta}_{\text{Adj.}}\right] = \frac{\sigma_\eta^2}{(T\sigma_x^2)} \tag{26}$$

where $\sigma_x^2 = T^{-1} \sum_{t=1}^{T} (x_t - \overline{x}_t)^2$ is the (biased) sample variance of x_t and σ_η^2 is the variance of the residuals from (16)

- Lewellen (2004) finds that bias-adjusted coefficients are similar to unadjusted coefficients, but that $\widehat{\beta}_{\text{Adj.}}$ has *much* lower variance and therefore strongly rejects the null of no predictability

Cochrane's (2008) defence

- Cochrane (2008) defends return predictability by directing attention to the inability of the log dividend-price to predict dividend growth. Consider the following system

$$\Delta d_{t+1} = \alpha_d + \beta_d (d_t - p_t) + \varepsilon_{d,t+1} \tag{27}$$
$$r_{t+1} = \alpha_r + \beta_r (d_t - p_t) + \varepsilon_{r,t+1} \tag{28}$$
$$d_{t+1} - p_{t+1} = \lambda + \rho (d_t - p_t) + \varepsilon_{dp,t+1} \tag{29}$$

- The loglinear approximation in (7) implies that the parameters are intimately linked

$$\beta_r = 1 + \beta_d - \phi\rho \tag{30}$$

- Suppose that $\phi = 0.97$ and we know that $\rho \leq 1$, then

$$\beta_r > \beta_d + 0.03 \tag{31}$$

- Cochrane (2008) finds that $\beta_d \approx 0$, providing us with indirect evidence that $\beta_r > 0$. That is, that stock returns are predictable!
- However, this is not the case around the world, where dividend predictability is more common (Engsted and Pedersen, 2010, Rangvid et al., 2014)

Multihorizon regressions

Long-horizon predictability

It may be relevant to forecast excess returns over a longer horizon than simply one-period ahead (e.g., for portfolio allocation decisions). In that case, we consider the multi-period counterpart to (2)

$$r_{t\to t+h} = \alpha_h + \beta_h x_t + \varepsilon_{t\to t+h} \tag{32}$$

where $r_{t\to t+h}$ denotes the log excess return (risk premia) from time t to $t+h$

- The h–period log excess return (risk premia) is computed as follows

$$r_{t\to t+h} = \ln\left(1 + r_{t\to t+h}\right) = \sum_{i=1}^{h} \ln\left(1 + r_{t+i}\right) \tag{33}$$

$$= \sum_{i=1}^{h} r_{t+i} \tag{34}$$

where r_{t+i} denotes the one-period log excess return (risk premia) earned from time $t+i-1$ to $t+i$

One- and multi-period predictability

- One- and multi-period predictability are intimately linked. Consider the one-period setting given by the system in (12) and (13) without constant terms for notational ease, then for $h = 2$ we have

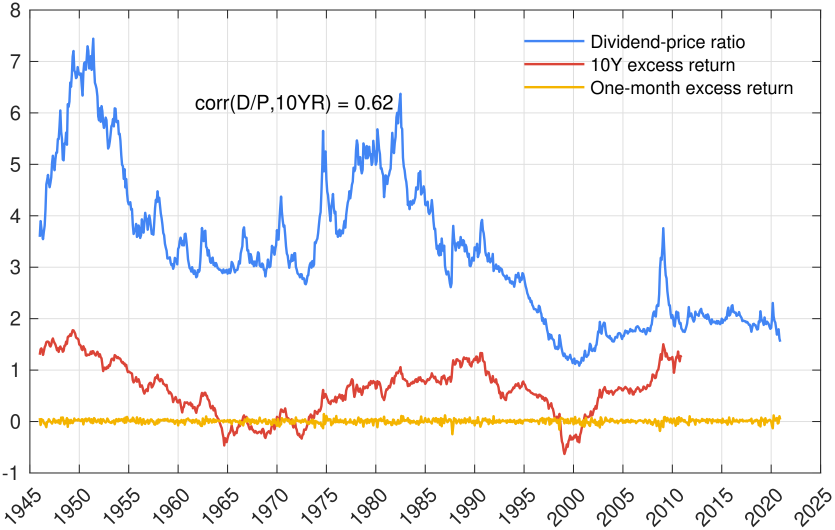
$$\begin{aligned} r_{t+1} + r_{t+2} &= \beta x_t + \varepsilon_{t+1} + \beta x_{t+1} + \varepsilon_{t+2} && (35) \\ &= \beta x_t + \varepsilon_{t+1} + \beta \rho x_t + \beta \nu_{t+1} + \varepsilon_{t+2} && (36) \\ &= \beta \left(1 + \rho\right) x_t + \beta \nu_{t+1} + \varepsilon_{t+1} + \varepsilon_{t+2} && (37) \end{aligned}$$

- Similarly, for $h = 3$ we obtain

$$r_{t+1} + r_{t+2} + r_{t+3} = \beta \left(1 + \rho + \rho^2\right) x_t + \text{error terms} \qquad (38)$$

- The key take-away is that we should expect β_h to increase in absolute size as a function of h if $\beta \neq 0$ and $\rho \neq 0$ up to $\beta_\infty = \beta / \left(1 - \rho\right)$

Illustrating long-horizon returns



Remarks on overlapping returns and standard errors

- In empirical studies of long-horizon predictability, we typically use overlapping data due to data limitations. In this case, $\varepsilon_{t\to t+h} \sim MA\,(h-1)$ by construction and usual OLS standard errors are no longer valid
- A natural solution to the problem of serially correlated errors is to use a HAC-type covariance estimator, e.g., the Newey and West (1987) estimator with a bandwidth (lag lenght) of $h-1$
- **Problem:** When the degree of time-overlap is large relative to the sample size (h/T is large), then the effective sample size is small $\Rightarrow$ Newey and West (1987) standard errors are poor approximations to the true standard errors
- Alternative estimators that perform better include Hansen and Hodrick (1980), Hodrick (1992), and Wei and Wright (2013).
- Ang and Bekaert (2007) show that Hodrick (1992) standard errors generally display better size (probability of rejecting a true null hypothesis) properties in long-horizon regression than Newey and West (1987) standard errors

Bootstrapping long-horizon regressions

- When using a (parametric) bootstrap to analyze predictability in long-horizon predictive regressions models, we use the one-period model as starting point and follow the same scheme as outlined above, but replace step 6 with

  6a. Use the T simulated one-period returns (or risk premia) r^*_{t+1} to build multi-period returns (or risk premia) $r^*_{t\to t+h}$

  6b. Estimate the multi-period model in (32) on the simulated data to obtain $\widetilde{\beta}^1_h$

- This approach does not require an estimate of the standard error and hence automatically deals with the overlapping data problem

- Another advantage of using a bootstrap is that we automatically and simultaneously account for small-sample bias

Out-of-sample predictability

Out-of-sample predictability

- Assessing predictability in-sample entails estimating the predictive regression using the full range of available observations

- Assessing out-of-sample predictability, conversely, entails using information available at time t only to forecast returns at time $t+1$

- To emulate a forecaster in real-time, we split the total sample (T) in two parts: in-sample (initial) and out-of-sample

$$\underbrace{t = 1,2,\ldots,R}_{\text{In-sample}},\underbrace{R+1,R+2,\ldots,T}_{\text{Out-of-sample}} \tag{39}$$

- One can either estimate the regression coefficients using a rolling or an expanding window of data (benefits/drawback?)

- Irrespective of choice, we end up with a sequence of forecasts $\{\widehat{r}_i\}_{i=R+1}^{T}$ and forecast errors $\{\widehat{\varepsilon}_i\}_{i=R+1}^{T}$ for evaluation

Statistical evaluation

- The natural benchmark is a historical average (HA) that assumes no predictability, i.e., constant expected excess returns (and it is *ridiculously* tough to beat!)

$$\bar{r}_{t+1} = \frac{1}{t} \sum_{i=1}^{t} r_i \tag{40}$$

- A popular statistical metric is the Mean Squared Forecast Error (MSFE)

$$\text{MSFE} = \frac{1}{T-R} \sum_{i=R+1}^{T} \left(r_i - \widehat{r}_i\right)^2 \tag{41}$$

- We can assess the degree of out-of-sample predictability using the out-of-sample R^2 (Fama and French, 1989, Campbell and Thompson, 2008)

$$\text{R}^2_{OS} = 1 - \frac{\text{MSFE}_x}{\text{MSFE}_{HA}} = 1 - \frac{\sum_{i=R+1}^{T} \left(r_i - \widehat{r}_i\right)^2}{\sum_{i=R+1}^{T} \left(r_i - \bar{r}_i\right)^2} \tag{42}$$

Interpreting the out-of-sample R^2

- The R^2_{OS} gives us the proportional reduction in MSFE for the predictive regression relative to the HA and is analogous to the in-sample R^2
- If $R^2_{OS} > 0$, then the predictive regression has lower average MSFE than the HA. That is, the predictor variable contains relevant information for forecasting r_{t+1} beyond what is already contained in the HA. Vice versa for $R^2_{OS} < 0$
- However, R^2_{OS} does not tell us whether the differences are large in a statistical sense. Suppose that we want to test

$$\mathcal{H}_0 : R^2_{OS} \leq 0 \quad \text{(No predictability)} \tag{43}$$

$$\mathcal{H}_1 : R^2_{OS} > 0 \tag{44}$$

- We consider two tests for this hypothesis. The Diebold and Mariano (1995) test and the Clark and West (2007) test

Diebold and Mariano (1995)

- The conventional approach is to use the Diebold and Mariano (1995) test for equal predictive accuracy (see also West (1996))
- To conduct the test, we first construct a time series of loss differentials

$$d_i = (r_i - \bar{r}_i)^2 - (r_i - \widehat{r}_i)^2 , \quad i = R+1, R+2, \ldots, T \tag{45}$$

- We can then test $\mathcal{H}_0 : \mathbb{E}[d_i] \leq 0$ by running the regression

$$d_i = \theta + \epsilon_i \tag{46}$$

  and perform a standard t-test on the constant θ using Newey and West (1987) standard errors to evaluate the null $\theta \leq 0$
- The test has a standard asymptotic distribution for non-nested models, but is severely undersized for nested models

Clark and West (2007)

- Clark and West (2007) propose an MSFE-adjusted test for nested models in which we first construct

$$f_i = (r_i - \bar{r}_i)^2 - \left[(r_i - \widehat{r}_i)^2 - (\bar{r}_i - \widehat{r}_i)^2\right], \quad i = R+1, R+2, \ldots, T \tag{47}$$

  where the last term adjust for the fact that we would expect the predictive regression to underperform because it has to estimate an additional parameter that is zero under the null hypothesis (i.e., due to noise)

- We can then test $\mathcal{H}_0 : \mathbb{E}\left[f_i\right] \leq 0$ by running the regression

$$f_i = \theta + \epsilon_i \tag{48}$$

  and perform a standard t-test on the constant θ using Newey and West (1987) standard errors to evaluate the null $\theta \leq 0$

- The test has (approximate) standard normal asymptotics for nested models, displays good small sample properties, and is a convenient/effective adjustment

Goyal and Welch's (2008) graphical device

Goyal and Welch's (2008) graphical device

Goyal and Welch (2008) propose a simple graphical device to assess predictability over time. Specifically, they propose to plot the cumulative difference in squared forecast errors (CDSFE)

$$\text{CDSFE}_t = \sum_{i=R+1}^{t} (r_i - \bar{r}_i)^2 - \sum_{i=R+1}^{t} (r_i - \widehat{r}_i)^2 , \quad t > i \tag{49}$$

- The CDSFE is a highly informative about the timing of the value of predictor information (if any) and is easy to interpret
  - A positive slope implies that the predictive models outperforms the benchmark in terms of MSFE
  - A negative slope implies that the predictive models underperforms the benchmark in terms of MSFE

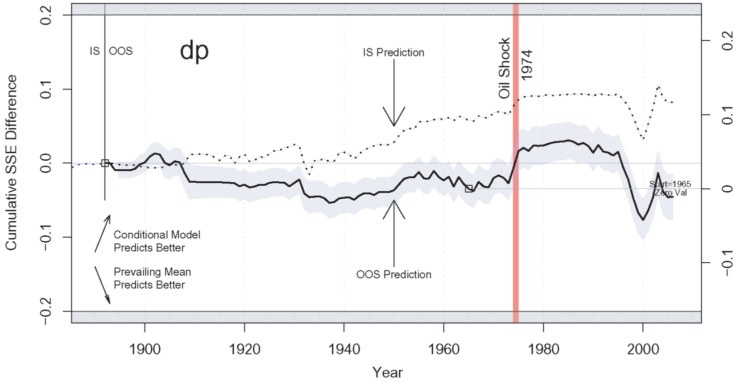
Goyal and Welch (2008)

- Goyal and Welch (2008) examine a long list of potential predictors and find limited evidence of in-sample predictability and essentially no evidence supporting out-of-sample predictability
- See also Goyal et al. (2021) for an updated version with newer predictors, but equally forceful conclusions about the lack of predictability

			Full Sample									1927–2005
				Forecasts begin 20 years after sample				Forecasts begin 1965				Sample
			IS	IS for	OOS			IS for	OOS			IS
	Variable	Data	$\overline{R}^2$	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power	$\overline{R}^2$
Full Sample, Not Significant IS												
dfy	Default yield spread	1919–2005	−1.18		−3.29	−0.14			−4.15	−0.12		−1.31
infl	Inflation	1919–2005	−1.00		−4.07	−0.20			−3.56	−0.08		−0.99
svar	Stock variance	1885–2005	−0.76		−27.14	−2.33			−2.44	+0.01		−1.32
d/e	Dividend payout ratio	1872–2005	−0.75		−4.33	−0.31			−4.99	−0.18		−1.24
lty	Long term yield	1919–2005	−0.63		−7.72	−0.47			−12.57	−0.76		−0.94
tms	Term spread	1920–2005	0.16		−2.42	−0.07			−2.96	−0.03		0.89
tbl	Treasury-bill rate	1920–2005	0.34		−3.37	−0.14			−4.90	−0.18		0.15
dfr	Default return spread	1926–2005	0.40		−2.16	−0.03			−2.82	−0.02		0.32
d/p	Dividend price ratio	1872–2005	0.49		−2.06	−0.11			−3.69	−0.09		1.67
d/y	Dividend yield	1872–2005	0.91		−1.93	−0.10			−6.68	−0.31		**2.71***
ltr	Long term return	1926–2005	0.99		−11.79	−0.76			−18.38	−1.18		0.92
e/p	Earning price ratio	1872–2005	1.08		−1.78	−0.08			−1.10	0.11		**3.20***

Goyal and Welch (2008)

- Goyal and Welch (2008) further examines out-of-sample predictability over time and argue that what limited evidence we might see is fully attributable to the 1974 oil crisis



Campbell and Thompson (2008)

- Campbell and Thompson (2008) defend out-of-sample predictability by showing that imposing a set of simple constraints substantially improves out-of-sample performance
- In particular, they impose that expected excess returns (risk premia) should be non-negative and that slope coefficients should align with theory. That is, they discipline their forecast as follows

$$\widehat{r}_{t+1} = \max\left\{0, \widehat{\alpha} + \max\left\{0, \widehat{\beta}\right\} x_t\right\} \tag{50}$$

- Clearly, one can impose either restriction separately or, as above, jointly. How they perform relative to an unrestricted forecast is an empirical question, but the empirical evidence points towards improvements

Campbell and Thompson (2008)

	Sample Begin	Forecast Begin	In-Sample t-statistic	In-Sample R-squared	Out-of-Sample R-squared with Different Constraints: Unconstrained	Positive Slope	Pos. Forecast	Both
A: Monthly Returns								
Dividend-price ratio	1872m2	1927m1	1.25	1.13%	−0.65%	0.05%	0.07%	0.08%
Earnings-price ratio	1872m2	1927m1	2.29	0.71	0.12	0.18	0.14	0.18
Smooth earnings-price ratio	1881m2	1927m1	1.85	1.36	0.33	0.42	0.38	0.43
Book-to-market	1926m6	1946m6	1.96	0.61	−0.43	−0.43	0.00	0.00
ROE	1936m6	1956m6	0.36	0.02	−0.93	−0.06	−0.93	−0.06
T-Bill rate	1920m1	1940m1	2.44	0.86	0.52	0.51	0.57	0.55
Long-term yield	1870m1	1927m1	1.46	0.19	−0.19	−0.19	0.20	0.20
Term spread	1920m1	1940m1	2.16	0.65	0.46	0.47	0.45	0.46
Default spread	1919m1	1939m1	0.74	0.10	−0.19	−0.19	−0.19	−0.19
Inflation	1871m5	1927m1	0.39	0.06	−0.22	−0.21	−0.18	−0.17
Net equity issuance	1927m12	1947m12	1.74	0.48	0.34	0.34	0.50	0.50
Consumption-wealth ratio	1951m12	1971m12	4.57	2.60	−1.36	−1.36	0.27	0.27
B: Annual Returns								
Dividend-price ratio	1872m2	1927m1	2.69	10.8	5.53	5.53	5.63	5.63
Earnings-price ratio	1872m2	1927m1	2.84	6.78	4.93	4.93	4.94	4.94
Smooth earnings-price ratio	1881m2	1927m1	3.01	13.57	7.89	7.89	7.85	7.85
Book-to-market	1926m6	1946m6	1.98	8.26	−3.38	−3.38	1.39	1.39
ROE	1936m6	1956m6	0.35	0.32	−8.60	−0.03	−8.35	−0.03
T-Bill rate	1920m1	1940m1	1.77	4.26	5.54	5.54	7.47	7.47
Long-term yield	1870m1	1927m1	0.91	0.77	−0.15	−0.15	2.26	2.26
Term spread	1920m1	1940m1	1.72	3.10	4.79	4.79	4.74	4.74
Default spread	1919m1	1939m1	0.07	0.01	−3.81	−3.81	−3.81	−3.81
Inflation	1871m5	1927m1	0.17	0.07	−0.71	−0.71	−0.71	−0.71
Net equity issuance	1927m12	1947m12	0.54	0.35	−4.27	−4.27	−2.38	−2.38
Consumption-wealth ratio	1951m12	1971m12	3.76	19.87	−7.75	−7.75	−1.48	−1.48

Remark on drawbacks of using predictors individually

- So far, we have mostly looked at predictors on their own. However, there are several drawbacks of such an approach
  1. It is difficult to identify the best predictor a priori
  2. Individual predictors are unstable and performs uneven over time
  3. Relying on individual predictors is therefore risky
- One can instead use, say, forecast combination (Timmermann, 2006, Rapach et al., 2010) to use information in multiple predictors without running into problems with overfitting in "kitchen sink" regressions
  - A simple approach is an equal-weighted $1/N$ strategy, but more exotic approaches exists (see, e.g., Rapach et al. (2010)). Whether the added estimation uncertainty improves on the $1/N$ strategy is an empirical question
  - Alternatives include dimension reduction techniques (principal components, partial least squares), shrinkage (ridge regression), variable selection methods (lasso, elastic net), or other machine learning techniques

Economic evaluation

Economic evaluation

- Investors may care more about economic than statistical value, and statistical criteria are not necessarily indicative of economic value (Leitch and Tanner, 1991, Marquering and Verbeek, 2004, Cenesizoglu and Timmermann, 2012)
- The idea is to equip the investor with a utility function and then compute utility gains for the predictive model over the benchmark
- This provides us with a direct measure of the economic value of return predictability that is closely tied to portfolio theory
  - Posit reasonable/convenient utility function for the investor (e.g., mean-variance utility)
  - Specify/derive an asset allocation rule based on return (risk premia) forecasts
  - Compare certainty equivalents (utility gains) for the investor when using predictive regression relative to benchmark

Mean-variance investor

- Consider the asset allocation problem of a risk-averse investor with mean-variance preferences and relative risk aversion γ of the form

$$\max_{\omega_t} \mathbb{E}_t \left[r_{p,t+1}\right] - \frac{1}{2}\gamma \mathrm{Var}_t \left[r_{p,t+1}\right] \tag{51}$$

- The investor chooses the weight ω_t to invest in the risky asset and the weight $(1-\omega_t)$ to invest in the risk-free rate using the Markowitz solution

$$\omega_t = \left(\frac{1}{\gamma}\right) \frac{\mathbb{E}_t \left[r_{t+1} - r_{f,t+1}\right]}{\mathrm{Var}_t \left[r_{t+1} - r_{f,t+1}\right]}, \tag{52}$$

  where $\mathbb{E}_t \left[r_{t+1} - r_{f,t+1}\right]$ is estimated using the predictive regression (or the benchmark model) and the variance is usually computed over a rolling window of realized excess returns

- The investor earns a realized out-of-sample portfolio return that depends on her portfolio allocations

$$r_{p,t+1} = (1-\omega_t)\, r_{f,t+1} + \omega_t r_{t+1} = r_{f,t+1} + \omega_t \left(r_{t+1} - r_{f,t+1}\right) \tag{53}$$

Certainty equivalent returns (utility gains)

- Given the asset allocation rule, we can compute the certainty equivalent return (CER) for the predictor as

$$\text{CER} = \mu_p - \frac{1}{2}\gamma\sigma_p^2 \tag{54}$$

where μ_p and σ_p^2 are the mean and variance of the resulting portfolio return

- We can then compute the annualized utility gain (for monthly data) as

$$\Delta = 1200 \times \left(\text{CER}_x - \text{CER}_{HA}\right) \tag{55}$$

which we can interpret as an annual portfolio management fee that the investor is willing to pay to access the information in the predictor variable

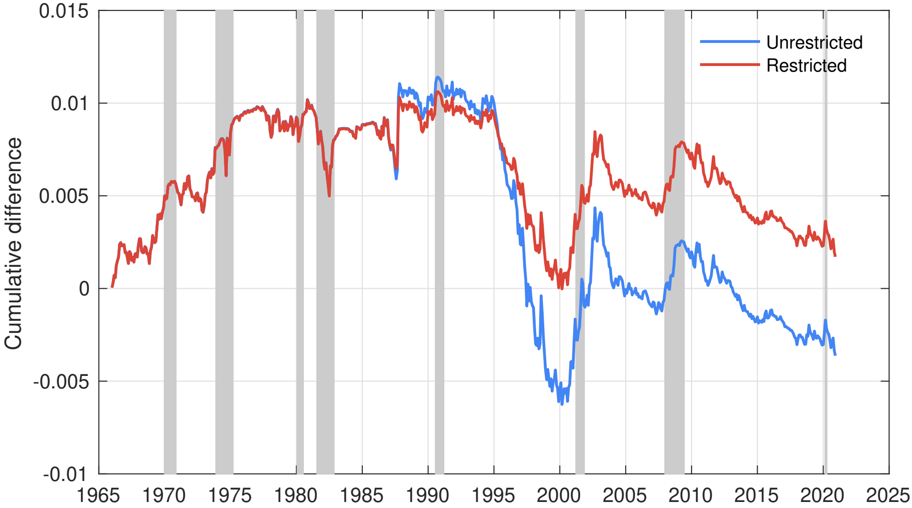
- We sometimes restrict the weights ω_t to lie between, say, $-0.5 \geq \omega_t \geq 1.5$ or similar to avoid extreme positions, i.e., we impose reasonable shorting and leverage constraints

Out-of-sample predictability using $d_t - p_t$

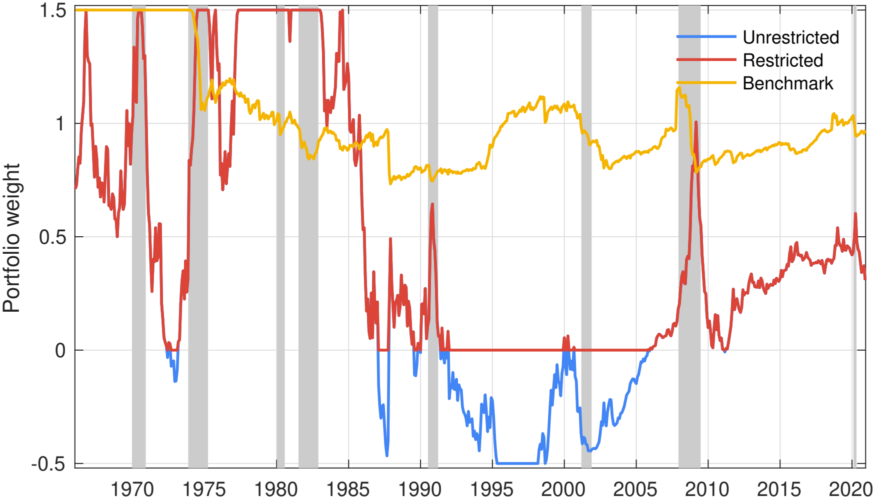
- Let us continue our example using the log dividend-price ratio to predict excess stock market returns in an out-of-sample setting
  - The initial window is $R = 240$ observations and we consider an expanding window forecasting scheme
  - The relative risk aversion is set to $\gamma = 3$ in the economic evaluation

	Unrestricted	Restricted
R^2_{OS}	-0.28	0.13
DM	[0.62]	[0.43]
CW	[0.10]	[0.06]
Δ	-0.66	0.00

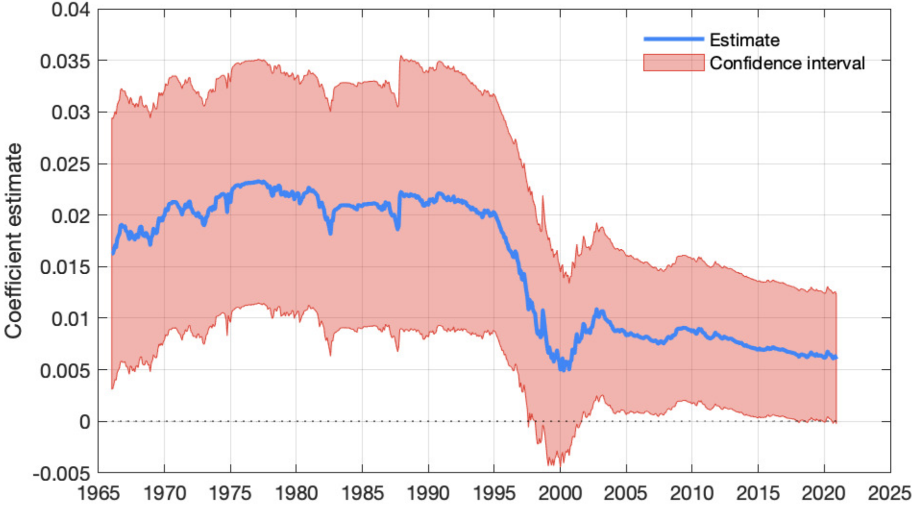
Goyal-Welch CDSFE plot



Mean-variance portfolio weights



Instability of the slope parameter



Time-varying predictability

Does predictability itself vary over time?

Can the mixed empirical evidence and resulting disagreement be caused by the degree of predictability itself varying over time?

- This addresses variation in predictability itself and the question becomes: *when (and if) does a given variable predict excess return*
- There is *ex post* empirical evidence that supports such an interpretation
  - **Stocks:** Henkel et al. (2011), Dangl and Halling (2012), Rapach et al. (2010), Rapach and Zhou (2013), and Farmer et al. (2021)
  - **Bonds:** Gargano et al. (2019), Andreasen et al. (2021), and Borup et al. (2021)
  - **Currencies:** Bacchetta and Van Wincoop (2004, 2013), Rossi (2013), and Fratzscher et al. (2015)

The dividend-price ratio as a classic example

- Consider the ongoing example of predicting excess stock market returns using the dividend-price ratio (Campbell and Shiller, 1988)

$$r_{t+1} = \alpha + \beta dp_t + \varepsilon_{t+1} \tag{56}$$

- Running an out-of-sample forecast exercise and partitioning *ex post* on recessions and expansions yields

	Overall		Expansions		Recessions	
	Unrestricted	Restricted	Unrestricted	Restricted	Unrestricted	Restricted
R^2_{OS}	-0.28	0.13	-1.08	-0.48	1.67	1.64
DM	[0.62]	[0.43]	[0.82]	[0.69]	[0.11]	[0.10]
CW	[0.10]	[0.06]	[0.26]	[0.18]	[0.06]	[0.05]
Δ	-0.66	0.00	-1.90	-1.13	7.43	7.35

- These are ex post (i.e., after the fact), but predictability may even itself be predictable ex ante (Borup et al., 2021)

Other asset classes

- Essentially everything above is universally applicable across asset classes such as Treasury bond markets, currency markets, and commodity markets. We are, in all cases, interested in predicting excess returns to some assets

$$r_{t+1} = \alpha + \boldsymbol{x}_t' \boldsymbol{\beta} + \varepsilon_{t+1} \tag{57}$$

- **Bonds:** In bonds, one would define the excess holding period return on a k-period bond as $rx_{t+\tau}^{(k)} = p_{t+\tau}^{(k-\tau)} - p_t^{(k)} + p_t^{(\tau)}$ and regress that on variables likely to influence bond risk premia (see, e.g., Fama and Bliss (1987), Cochrane and Piazzesi (2005), Cooper and Priestley (2009), Ludvigson and Ng (2009), Cieslak and Povala (2015), Eriksen (2017), and Bauer and Hamilton (2018))

Currencies

- Remember, the completeness condition for defining exchange rates:

$$\frac{S_{t+1}}{S_t} = \frac{\tilde{M}_{t+1}}{M_{t+1}}, \tag{58}$$

- Consider a continuous-time framework, such that

$$dM_t = -r_{f,t} M_t dt - \lambda_t M_t dW_t, \tag{59}$$

$$d\tilde{M}_t = -\tilde{r}_{f,t} \tilde{M}_t dt - \tilde{\lambda}_t \tilde{M}_t dW_t, \tag{60}$$

$$\tag{61}$$

- Applying Ito's lemma to log exchange rate

$$ds_t = (r_{f,t} - \tilde{r}_{f,t} - \frac{1}{2}(\tilde{\lambda}_t{}' \tilde{\lambda}_t - \lambda_t' \lambda_t))dt + (\tilde{\lambda}_t - \lambda_t) dW_t \tag{62}$$

Currencies, cont.

- A simple Euler-discretization of the dynamics of the log exchange rate reveals

$$\mathbb{E}(\Delta s_{t+1}) \approx (r_{f,t} - \tilde{r}_{f,t} - \frac{1}{2}(\tilde{\lambda}_t{}'\tilde{\lambda}_t - \lambda_t'\lambda_t)) \tag{63}$$

- Expected short-term exchange-rate movements can be decomposed into two components:
  - The difference in interest rates
  - A function of the difference in market-price of risk (risk compensation per unit of risk)
    ➜ If risk compensation is different between two countries, the exchange rate must reflect that!

Currency predictors

- Traditionally, the literature has focused on macroeconomic fundamentals (see, e.g., Meese and Rogoff (1983a), Fama (1984), Engel and West (2005), Della Corte et al. (2009), Rossi (2013), and Engel (2014))
- But you can naturally also focus on TS predictability of currency factors, for instance:
    - Carry trades: Commodity returns (Bakshi and Panayotov, 2013a, Ready et al., 2017), Global variance innovations (Bakshi and Panayotov, 2013a), Credit risk (Della Corte et al., 2021), Liquidity related (such as the TED spread) (Brunnermeier et al., 2008)
    - Dollar factor: average forward discount (Lustig et al., 2014), growth in commercial papers (Fang and Liu, 2021), Variance risk premia imbalances (Kjær and Posselt, 2022), CIP violations (Jiang et al., 2021)

A small remark

- Everything we have done so far has been motivated theoretically. Meaning that the stochastic discount factor is a somewhat exogenous process which we then try to model
- ... but at the end-of-the-day a trade occurs because a buyer and a seller agree on a price
- A more recent literature focus on contraints on intermediaries impact asset pricing
- For inspiration, see among others: Haddad and Muir (2021), Fang and Liu (2021), Haddad and Sraer (2020), and He and Krishnamurthy (2018) for a theoretical survey

Back to the SDF

Two equations for returns

- We have now seen two simple linear expressions for returns:

$$r_{t+1} = \beta' f_{t+1} + \eta_{t+1} \tag{64}$$
$$r_{t+1} = b_0 + x_t b_x + \varepsilon_{t+1} \tag{65}$$
$$\tag{66}$$

- It seems natural to ask whether the two equations are consistent!
- In other words, is the predictabive ability of x_t consistent with the linear factor model spanned by f_t?
- Let us examine that! (you can read more in Kirby (1998) and Bakshi and Panayotov (2013b))

Setup

- Consider a linear SDF, which naturally satisfies

$$\mathbb{E}_t(r_{t+1}M_{t+1}) = 0 \quad (67)$$
$$M_{t+1} = 1 - \tilde{b}' f_{t+1} \quad (68)$$

- For simplicity, assume $\mathbb{E}(f_{t+1}) = 0$ and define $x_t^* = [1\ x_t]$, $b = [b_0\ b_x]'$
- The predictive regression in (3) for asset k can be written as

$$r_{t+1}^{(k)} = x_t^* b^u + \varepsilon_{k,t+1} \quad (69)$$

with $b^u = (\mathbb{E}(x_t^{*\prime} x_t^*))^{-1}\mathbb{E}(r_{t+1}^{(k)} x_t^{\prime *})$. The superscript, u, is due to the model is unrestricted. This will make sense in the next slide...

Restriction on the predictive model

- Combining the predictive model with the pricing model, imply that the predictive coefficient b must be

$$b^r = -\mathbb{E}(x_t^{*\prime} x_t^*)^{-1} cov(M_{t+1}, r_{t+1}^{(k)} x_t^{*\prime}) \tag{70}$$

- Where superscript r is due to the model is restricted by the assumed SDF.
- So, we have two expression for the same coefficient. Wouldn't it be nice to test for whether these expressions are consistent...
- Can you smell a GMM framework coming up?

A test for consistency between SDF and predictive model

- We can jointly estimate the parameters, using the following moment conditions
$$\mathbb{E}\begin{pmatrix} (r_{t+1}^{(k)} - \beta' f_{t+1}) \otimes f_{t+1} \\ (r_{t+1}^{(k)} - x_t^* b^u) \otimes x_t^* \\ (\beta' f_{t+1} - x_t^* b^r) \otimes x_t^* \end{pmatrix} = 0 \tag{71}$$
- We can test the restriction implied by defining $b^* = b^u - b^r$ and test for b^* being a zero vector (Wald test). The system is just-identified
- Let us test whether CAPM is consistent with predictability embedded in the dividend-price ratio!

Empirical example

- Let us consider whether the predictive ability of the variance risk premium is consistent with CAPM, see the livescript *Kirby_VP_CAPM.mlx*.

Potential projects

Potential projects

- Reexamine in-sample predictability across multiple predictors, horizons, and/or countries while potentially accounting for small sample bias
- Reexamine out-of-sample predictability across multiple predictors and/or countries using both statistical and economic evaluations
- Propose and evaluate a *new* predictor in-sample (with bootstrapping) and/or out-of-sample from a statistical and economic perspective
- Examine time-varying (state-dependent) predictability for a set of existing predictors using existing or new state-variables
- Examine time-varying (state-dependent) predictability across different asset classes and/or countries
- Combine the above with machine learning techniques and/or forecast combination and dimension reduction techniques

References

Andreasen, M. M., T. Engsted, S. V. Møller, and M. Sander (2021). The yield spread and bond return predictability in expansion and recessions. *Review of Financial Studies 34*(6), 2773–2812.

Ang, A. and G. Bekaert (2007). Stock return predictability: Is it there? *Review of Financial Studies 20*(3), 651–707.

Atanasov, V., S. V. Møller, and R. Priestley (2020). Consumption fluctuations and expected returns. *Journal of Finance 75*(3), 1677–1713.

Bacchetta, P. and E. Van Wincoop (2004). A scapegoat model of exchange-rate fluctuations. *American Economic Review 94*, 114–118.

Bacchetta, P. and E. Van Wincoop (2013). On the unstable relationship between exchange rates and macroeconomic fundamentals. *Journal of International Economics 91*, 18–26.

Bakshi, G. and G. Panayotov (2013a). Predictability of currency carry trades and asset pricing implications. *Journal of financial economics 110*(1), 139–163.

Bakshi, G. and G. Panayotov (2013b). Predictability of currency carry trades and asset pricing implications. *Journal of Financial Economics 110*(1), 139–163.

Bauer, M. D. and J. D. Hamilton (2018). Robust bond risk premia. *Review of Financial Studies 31*(2), 399–448.

Bianchi, D., M. Büchner, and A. Tamoni (2021). Bond risk premia with machine learning. *Review of Financial Studies 34*(2), 1046–1089.

Borup, D., J. N. Eriksen, M. M. Kjær, and M. Thyrsgaard (2021). Predicting bond return predictability. Working paper, Aarhus University.

Brunnermeier, M. K., S. Nagel, and L. H. Pedersen (2008). Carry trades and currency crashes. *NBER macroeconomics annual 23*(1), 313–348.

Campbell, J. Y. and R. J. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies 1*(3), 195–228.

Campbell, J. Y. and R. J. Shiller (1991). Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies 58*(3), 495–514.

Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies 21*(4), 1509–1531.

Cenesizoglu, T. and A. Timmermann (2012). Do return prediction models add economic value? *Journal of Banking & Finance 36*(11), 2974–2987.

Cheung, Y.-W., M. D. Chinn, A. G. Pascual, and Y. Zhang (2019). Exchange rate prediction redux: New models, new data, and new currencies. *Journal of International Money and Finance 95*, 332–362.

Cieslak, A. and P. Povala (2015). Expected returns in Treasury bonds. *Review of Financial Studies 28*(10), 2859–2901.

Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics 138*, 291–311.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies 21*(4), 1533–1575.

Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. *American Economic Review 95*(1), 138–160.

Cooper, I. and R. Priestley (2009). Time-varying risk premiums and the output gap. *Review of Financial Studies 22*(7), 2801–2833.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106*, 157–181.

References

Della Corte, P., L. Sarno, M. Schmeling, and C. Wagner (2021). Exchange rates and sovereign risk. *Management Science*.

Della Corte, P., L. Sarno, and I. Tsiakas (2009). An economic evaluation of empirical exchange rate models. *Review of Financial Studies 22*(9), 3491–3530.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 134–144.

Engel, C. (2014). Exchange rates and interest parity. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4, Chapter 8, pp. 453–522. Elsevier.

Engel, C. and K. D. West (2005). Exchange rates and fundamentals. *Journal of Political Economy 113*(3), 485–517.

Engsted, T. and T. Q. Pedersen (2010). The dividend-price ratio does predict dividend growth: International evidence. *Journal of Empirical Finance 17*(4), 585–605.

Eriksen, J. N. (2017). Expected business conditions and bond risk premia. *Journal of Financial and Quantitative Analysis 52*(4), 1667–1703.

Fama, E. F. (1984). Forward and spot exchange rates. *Journal of Monetary Economics 14*, 319–338.

Fama, E. F. and R. R. Bliss (1987). The information in long-maturity forward rates. *American Economic Review 77*(4), 680–692.

Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics 22*, 3–25.

Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics 25*, 23–49.

Fang, X. and Y. Liu (2021). Volatility, intermediaries, and exchange rates. *Journal of Financial Economics 141*(1), 217–233.

Farmer, L. E., L. Schmidt, and A. Timmermann (2021). Pockets of predictability. Working paper.

Fratzscher, M., D. Rime, L. Sarno, and G. Zinna (2015). The scapegoat theory of exchange rates: The first tests. *Journal of Monetary Economics 70*, 1–21.

Gargano, A., D. Pettenuzzo, and A. Timmermann (2019). Bond return predictability: Economic value and links to the macroeconomy. *Management Science 65*(2), 508–540.

Ghysels, E., C. Horan, and E. Moench (2018). Forecasting through the rearview mirror: Data revisions and bond return predictability. *Review of Financial Studies 31*(2), 678–714.

Goyal, A. and I. Welch (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

Goyal, A., I. Welch, and A. Zafirov (2021). A comprehensive look at the empirical performance of equity premium prediction ii. Swiss Finance Institute Research Paper Series No. 21-85.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Haddad, V. and T. Muir (2021). Do intermediaries matter for aggregate asset prices? *The Journal of Finance 76*(6), 2719–2761.

Haddad, V. and D. Sraer (2020). The banking view of bond risk premia. *The Journal of Finance 75*(5), 2465–2502.

Hansen, L. P. and R. J. Hodrick (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy 88*(5), 829–853.

References

He, Z. and A. Krishnamurthy (2018). Intermediary asset pricing and the financial crisis. *Annual Review of Financial Economics 10*, 173–197.

Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of Financial Economics 99*, 560–580.

Hodrick, R. J. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies 5*(3), 357–386.

Jiang, Z., A. Krishnamurthy, and H. Lustig (2021). Foreign safe asset demand and the dollar exchange rate. *The Journal of Finance 76*(3), 1049–1089.

Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika 41*(3-4), 403–404.

Kirby, C. (1998). The restrictions on predictability implied by rational asset pricing models. *Review of Financial Studies 11*(2), 343–382.

Kjær, M. M. and A. M. Posselt (2022). The us dollar and variance risk premia imbalances. *Available at SSRN 4078021*.

Kothari, S. P. and J. Shanken (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics 44*, 169–203.

Leitch, G. and J. E. Tanner (1991). Economic forecast evaluation: Profits versus the conventional error measures. *American Economic Review 81*(3), 580–590.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics 74*, 209–235.

Li, J., I. Tsiakas, and W. Wang (2015). Predicting exchange rates out of sample: Can economic fundamentals beat the random walk? *Journal of Financial Econometrics 13*(2), 293–341.

Ludvigson, S. C. and S. Ng (2009). Macro factors in bond risk premia. *Review of Financial Studies 22*(12), 5027–5067.

Lustig, H., N. Roussanov, and A. Verdelhan (2014). Countercyclical currency risk premia. *Journal of Financial Economics 111*(3), 527–553.

Marquering, W. and M. Verbeek (2004). The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis 39*(2), 407–429.

Meese, R. A. and K. Rogoff (1983a). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics 14*, 3–24.

Meese, R. A. and K. Rogoff (1983b). The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification? In J. A. Frenkel (Ed.), *Exchange Rates and International Macroeconomics*, pp. 67–112. University of Chicago Press.

Molodtsova, T. and D. H. Papel (2009). Out-of-sample exchange rate predictability with Taylor rule fundamentals. *Journal of International Economics 77*, 167–180.

Nelson, C. R. and M. J. Kim (1993). Predictable stock returns: The role of small sample bias. *Journal of Finance 48*(2), 641–661.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Rangvid, J., M. Schmeling, and A. Schrimpf (2014). Dividend predictability around the world. *Journal of Financial and Quantitative Analysis 49*(5/6), 1255–1277.

Rapach, D. and G. Zhou (2013). Forecasting stock returns. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Chapter 6, pp. 328–383. Elsevier.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies 23*(2), 821–862.

References

Ready, R., N. Roussanov, and C. Ward (2017). Commodity trade and the carry trade: A tale of two countries. *The Journal of Finance 72*(6), 2629–2684.

Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature 51*(4), 1063–1119.

Sarno, L., P. Schneider, and C. Wagner (2016). The economic value of predicting bond risk premia. *Journal of Empirical Finance 37*, 247–267.

Stambaugh, R. F. (1986). Bias is regressions with lagged stochastic regressors. Unpublished manuscript, University of Chicago.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*, 375–421.

Thornton, D. L. and G. Valente (2012). Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. *The Review of Financial Studies 25*(10), 3141–3168.

Timmermann, A. (2006). Forecast combination. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, Chapter 4, pp. 135–196. Elsevier.

Wei, M. and J. H. Wright (2013). Reverse regressions and long-horizon forecasting. *Journal of Applied Econometrics 28*, 353–371.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica 64*(5), 1067–1084.

Portfolio Sorting

Empirical Asset Pricing

Mads Markvart Kjær

Department of Economics and Business Economics, Aarhus University

E-mail: mads.markvart@econ.au.dk

Spring 2023

Back to the pricing equation

- Remember;

$$\mathbb{E}_t(R_{i,t+1}) - R_{f,t} = -\frac{\text{Cov}_t(M_{t+1}, R_{i,t+1})}{Var_t(M_{t+1})}\frac{Var_t(M_{t+1})}{\mathbb{E}_t(M_{t+1})} \quad (1)$$

- Meaning that expected returns are linear in SDF exposure!
- ... But what if we hypotesize that a variable f_t belongs to the SDF and we have absolute no clue on the functional form?

  ➜ We can apply a portfolio sort!

,

Why do we care?

- Portfolio sorts are central to the empirical asset pricing literature and is a commonly applied methodology
- Can some characteristic of the assets explain cross-sectional variation?
- Can good performing stocks be identified and at which cost?

Outcome of lecture

After the lecture, you should have

- knowlegde and understading of
  - Portfolio sorting based on characteristics and covariances, the construction of zero-cost risk factors, and their implications for market efficiency
- and be able to
  - Discuss and conduct a portfolio sort using individual assets, evaluate the resulting portfolio returns, construct long-short factors, evaluate their returns, and reflect on the implications

The King of Univeriate Portfolio sort!

TURAN G. BALI

- Robert Parker Chair Professor of Finance
  McDonough School of Business
  Georgetown University
- Associate Editor
  - Journal of Financial and Quantitative Analysis
  - Management Science

A world of oppotunities!

Journal of Financial Economics 131 (2019) 619–642

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

ELSEVIER

Common risk factors in the cross-section of corporate bond returns☆

Jennie Bai, Turan G. Bali, Quan Wen

McDonough School of Business, Georgetown University, 3700 O St., Washington, DC, 20057, USA

Since the empirical distribution of bond returns is skewed, peaked around the mode, and has fat tails, downside risk—defined as a nonlinear function of volatility, skewness, and kurtosis—is expected to play a major role in the cross-sectional pricing of corporate bonds.

4.2. Univariate portfolio analysis

We first examine the significance of a cross-sectional relation between VaR and future corporate bond returns using portfolio-level analysis. For each month from July

A world of oppotunities!

Journal of Financial Economics 135 (2020) 725–753

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

ELSEVIER

Left-tail momentum: Underreaction to bad news, costly arbitrage and equity returns ☆

Check for updates

Yigit Atilgan[a], Turan G. Bali[b,*], K. Ozgur Demirtas[a], A. Doruk Gunaydin[a]

[a] Sabanci University, School of Management, Orhanli Tuzla 34956, Istanbul, Turkey
[b] Georgetown University, McDonough School of Business, Washington, D.C. 20057, USA

ket betas, lower book-to-market ratios, and lower idiosyncratic volatilities. Finally, there is a highly significant, positive correlation between idiosyncratic volatility and lottery demand.

3.2. *Univariate portfolio analysis*

In this section, we perform univariate portfolio-level analysis, where deciles are formed every month by sorting stocks based on their value-at-risk metrics at the 1% level and one-month-ahead returns are calculated for each

A world of oppotunities!

The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXIX, NO. 5 • OCTOBER 2014

The Joint Cross Section of Stocks and Options

BYEONG-JE AN, ANDREW ANG, TURAN G. BALI, and NUSRET CAKICI*

ABSTRACT

Stocks with large increases in call (put) implied volatilities over the previous month tend to have high (low) future returns. Sorting stocks ranked into decile portfolios by past call implied volatilities produces spreads in average returns of approximately 1% per month, and the return differences persist up to six months. The cross section of stock returns also predicts option implied volatilities, with stocks with high past returns tending to have call and put option contracts that exhibit increases in implied volatility over the next month, but with decreasing realized volatility. These predictability patterns are consistent with rational models of informed trading.

A world of oppotunities!

Growth Options and Related Stock Market Anomalies: Profitability, Distress, Lotteryness, and Volatility

Turan G. Bali, Luca Del Viva, Neophytos Lambertides, and Lenos Trigeorgis*

including the FISKEW factor whenever the value-minus-growth return spread appears significant in parts of the sample period. This holds using either the decile 10 minus decile 1 return spread on the book-to-market portfolios or the HML factor of Fama and French (1993). We provide a discussion and corresponding results in Section II and Figures A.1 and A.2 of the Supplementary Material.

D. Univariate Portfolio Analysis and Economic Significance

Table 5 provides further evidence concerning the economic significance of our growth-option-driven skewness measure, $E[ISKEW]_{GO}$, based on univariate portfolios. For each month, we form equal-weighted (EW) and value-weighted (VW) decile portfolios by sorting individual stocks based on their growth-option-driven expected idiosyncratic skewness, $E[ISKEW]_{GO}$, where decile 1 contains stocks with the lowest $E[ISKEW]_{GO}$ and decile 10 contains stocks with the highest $E[ISKEW]_{GO}$. Table 5 reports, by row, the average $E[ISKEW]_{GO}$, the average

A world of oppotunities!

informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 64, No. 9, September 2018, pp. 4137–4155
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Unusual News Flow and the Cross Section of Stock Returns

Turan G. Bali,[a] Andriy Bodnaruk,[b] Anna Scherbina,[c,d] Yi Tang[e]

[a] McDonough School of Business, Georgetown University, Washington, DC 20057; [b] College of Business Administration, University of Illinois at Chicago, Chicago, Illinois 60607; [c] Graduate School of Management, University of California at Davis, Davis, California 95616; [d] International Business School, Brandeis University, Waltham, Massachusetts 02453; [e] Gabelli School of Business, Fordham University, New York, New York 10023

Contact: turan.bali@georgetown.edu (TGB); bodnaruk@uic.edu (AB); ascherbina@ucdavis.edu (AS); ytang@fordham.edu (YT)

2.2. Volatility Shocks and Future Returns

In the remainder of the section, we establish a robust negative relation that exists between the current month's volatility shocks and the following month's returns.

2.2.1. Volatility-Shock-Sorted Portfolios. To establish the negative predictive ability of volatility shocks on future returns, every month we sort stocks into decile portfolios based on $IVOL^{shock}$. We then calculate next month's equal- and value-weighted portfolio returns and the return differentials between

A world of oppotunities!

informs.
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 63, No. 11, November 2017, pp. 3760–3779
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Dynamic Conditional Beta Is Alive and Well in the Cross Section of Daily Stock Returns

Turan G. Bali,[a] Robert F. Engle,[b] Yi Tang[c]

[a] McDonough School of Business, Georgetown University, Washington, DC 20057; [b] Stern School of Business, New York University, New York, New York 10012; [c] School of Business, Fordham University, New York, New York 10023
Contact: turan.bali@georgetown.edu (TGB); rengle@stern.nyu.edu (RFE); ytang@fordham.edu (YT)

3. Market Beta and the Cross Section of Daily Returns

This section examines the significance of a cross-sectional relation between the unconditional beta, the dynamic conditional beta, and daily stock returns based on the long-short equity portfolios. First, we perform univariate portfolio-level analysis for the unconditional measures of market beta. Second, we test the predictive power of the DCC beta based on univariate portfolio-level analysis. Finally, we provide average portfolio characteristics of the DCC beta–sorted port-

A world of oppotunities!

Liquidity Shocks and Stock Market Reactions

Turan G. Bali
McDonough School of Business, Georgetown University

Lin Peng
Zicklin School of Business, Baruch College, City University of New York

Yannan Shen
McCallum Graduate School of Business, Bentley University

Yi Tang
Schools of Business, Fordham University

2. Cross-Sectional Relation Between Liquidity Shocks and Stock Returns

The significantly positive correlation between liquidity shocks and one-month-ahead stock returns suggests that negative liquidity shocks (reductions in liquidity) are related to lower future stock returns, and vice versa. In this section, we perform formal analysis, and show that the pricing effect documented in this paper cannot be explained by other risk factors and stock characteristics that are known to predict future stock returns in the cross-section.

2.1 Univariate portfolio-level analysis

We begin our empirical analysis with univariate portfolio sorts. For each month, we sort common stocks trading on NYSE/AMEX/NASDAQ into decile

A world of oppotunities!

Hybrid Tail Risk and Expected Stock Returns: When Does the Tail Wag the Dog?

Turan G. Bali
Georgetown University

Nusret Cakici
Fordham University

Robert F. Whitelaw
New York University and NBER

3. Preliminary Evidence

Given the number of potential control variables, that is, other stock characteristics that may influence returns, the Fama-MacBeth cross-sectional regression approach may be the natural way to examine the predictive power of measures of tail risk. We turn to these regressions in Section 4; however, to get an initial feel for the data, we first look at univariate sorts on the basis of our three tail risk measures and the associated characteristics of the portfolios.

3.1 Average returns for univariate portfolio sorts

Table 1 presents the average monthly returns for the equal-weighted and value-weighted decile portfolios that are formed by sorting the NYSE, AMEX, and NASDAQ stocks based on our three tail risk measures—

A world of oppotunities!

Journal of Financial Economics 126 (2017) 471–489

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

ELSEVIER

Is economic uncertainty priced in the cross-section of stock returns?☆

CrossMark

Turan G. Bali[a,*], Stephen J. Brown[b,c], Yi Tang[d]

[a] McDonough School of Business, Georgetown University, 37th and O Streets, N.W., Washington, D.C. 20057, USA
[b] Monash Business School, Monash University, 900 Dandenong Road, Caulfield East, Victoria 3145, Australia
[c] Stern School of Business, New York University, 44 West fourth Street, New York, NY 10012, USA
[d] Gabelli School of Business, Fordham University, 45 Columbus Avenue, New York, NY 10023, USA

4. Empirical results

In this section, we conduct parametric and nonparametric tests to assess the predictive power of the uncertainty beta over future stock returns. First, we start with univariate portfolio-level analyses. Second, we discuss average stock characteristics to obtain a clear picture of the composition of the uncertainty beta portfolios. Third, we conduct bivariate portfolio-level analyses to examine the predictive power of the uncertainty beta after controlling for

A world of oppotunities!

JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 56, No. 5, Aug. 2021, pp. 1653–1678

doi:10.1017/S0022109020000538

The Macroeconomic Uncertainty Premium in the Corporate Bond Market

Turan G. Bali
Georgetown University McDonough School of Business
Turan.Bali@georgetown.edu (corresponding author)

Avanidhar Subrahmanyam
University of California at Los Angeles Anderson Graduate School of Management
subra@anderson.ucla.edu

Quan Wen
Georgetown University McDonough School of Business
Quan.Wen@georgetown.edu

Abstract

III. Empirical Results

In this section, we conduct parametric and nonparametric tests to assess the predictive power of the uncertainty beta over future corporate bond returns. We start with univariate portfolio-level analyses, presenting the average returns, alphas, and average bond characteristics of β^{UNC}-sorted portfolios. Second, we conduct bivariate portfolio-level analyses to examine the predictive power of the uncertainty beta after controlling for well-known measures of systematic risk, liquidity, and bond characteristics. Third, we provide an alternative risk-based explanation of the

A world of oppotunities!

Journal of Financial Economics 99 (2011) 427–446

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

ELSEVIER

Maxing out: Stocks as lotteries and the cross-section of expected returns ☆

Turan G. Bali [a,1], Nusret Cakici [b,2], Robert F. Whitelaw [c,d,*]

[a] Department of Economics and Finance, Zicklin School of Business, Baruch College, One Bernard Baruch Way, Box 10-225, New York, NY 10010, United States
[b] Department of Finance, Fordham University, Fordham University, 113 West 60th Street, New York, NY 10023, United States
[c] Stern School of Business, New York University, 44 W. 4th Street, Suite 9-190, New York, NY 10012, United States
[d] NBER, United States

2.2. *Univariate portfolio-level analysis*

Table 1 presents the value-weighted and equal-weighted average monthly returns of decile portfolios that are formed by sorting the NYSE/Amex/Nasdaq stocks based on the maximum daily return within the previous month (MAX). The results are reported for the sample period July 1962–December 2005.

Portfolio 1 (low MAX) is the portfolio of stocks with the lowest maximum daily returns during the past month,

Portfolio sorts

- Portfolio sorts are useful in identifying and assessing variables that can predict cross-sectional variation in future returns
- Typical sorting variables are *characteristic* such as size, value, CAPM-β, idiosyncratic volatility, or past returns of the assets
- But the sorting variable can also be exposure to economically motivated risk factors, e.g., market-wide volatility or macroeconomic risks
- Once sorted, we are interested in;
  - Assessing the cross-sectional relation between the sorting variable(s) and average returns
  - The returns to a zero-cost long-short (spread) portfolio

Advantages

- Portfolio sorts offer several advantages as a methodology

    1. Portfolio sorts does not require any a priori assumptions about the cross-sectional relationship between the sorting variable and expected returns
    2. Sorting stocks into portfolios can assist in the discovery of non-linear and linear cross-sectional relations alike
    3. It is highly flexibly and lets the researcher control much of the setup

Disadvantages

- On the other side, there are also a few drawbacks with the approach

  1. It is only possible to control for a very limited set of factors when examining the cross-sectional relations of interest
  2. The univariate (bivariate) approach only considers one (two) sorting variable(s) without controlling for other potentially important factors
  3. Many choices are left to the researcher without any clear guidance on *correct* implementations (e.g., data-snooping concerns)

Univariate Portfolio Sort

Univariate portfolio sort

- A univariate portfolio sort considers a single sorting variable $F_{i,t}$ for each security $i = 1,2,\ldots,N$ at time $t = 1,2,\ldots,T$
- The objective is to study the cross-sectional relationship between the factor $F_{i,t}$ and the future (excess) return to individual assets

Steps in univariate portfolio sorting

We can distill the process of conducting and interpreting a univariate portfolio sort into four basic steps

1. Calculate breakpoints for dividing the universe of assets into portfolios
2. Allocate assets into portfolios using the breakpoints
3. Compute portfolio returns in a meaningful way
4. Examine the cross-sectional variation in average portfolio (excess) returns

Step 1: Computing breakpoints

- The first step is to compute breakpoints for the cross-sectional distribution of the sorting factor $F_{i,t}$ for each time period t
- Suppose that we wish to form n_p portfolios, then we need $n_p - 1$ breakpoints for portfolio formation
- Let p_k denote the kth percentile of the values of $F_{i,t}$ across all available assets and denote by $\mathcal{B}_{k,t}$ the kth breakpoint at time t, then

$$\mathcal{B}_{k,t} = \text{Percentile}_{p_k}\left(F_{i,t}\right) \tag{2}$$

- The percentiles, and by extension the breakpoints, are increasing in k so that $0 < p_1 < p_2 < \cdots < p_{n_p - 1}$ and $\mathcal{B}_{1,t} \leq \mathcal{B}_{2,t} \leq \cdots \leq \mathcal{B}_{n_p - 1}$ for all t

Choices in breakpoint determination

Choices in breakpoint determination

There are three key choices left to the researcher in determining the breakpoints

1. **Choosing the assets for the breakpoints:** We can determine the breakpoints using all assets or a subset of the assets. As an example, one can use all available stocks on the CRSP tape or only NYSE stocks
2. **Choosing the number of portfolios:** This choice is largely a trade-off between the number of assets in each portfolio and the cross-sectional variation in expected returns that can reliably be identified using the sorting factor $F_{i,t}$
3. **Choosing percentiles:** The percentiles can be evenly spaced or unevenly spaced. Fama and French (1993), as en example, use the 30th and 70th percentiles

- In the end, we need to make economically motivated and well-argued choices for these parameters that suit the study at hand

Step 2: Portfolio formation

- Suppose that we have $k = 1,2,\ldots,n_p - 1$ breakpoints defined using the chosen percentiles, and define $\mathcal{B}_{0,t} = -\infty$ and $\mathcal{B}_{n_p,t} = \infty$ to exhaust all possible values of the sorting variable $F_{i,t}$

- We can then identify all securities i that belong to the kth portfolio formed at time t as the set of securities with values of $F_{i,t}$ that satisfy the relation

$$P_{k,t} = \left\{ i \middle| \mathcal{B}_{k-1,t} \leq F_{i,t} \leq \mathcal{B}_{k,t} \right\}, \qquad k = 1,2,\ldots,n_p \tag{3}$$

- Note that this approach puts all securities with the lowest values of the sorting factor $F_{i,t}$ in the first portfolio and all securities with the largest values of the sorting factor $F_{i,t}$ in the last portfolio by construction

Step 3: Computing portfolio returns

- Let $N_{k,t}$ denote the number of securities in portfolio k at time t, then equal-weighted portfolio returns for portfolio k are computed as

$$r_{k,t} = \frac{1}{N_{k,t}} \sum_{i=1}^{N_{k,t}} r_{i,t} \tag{4}$$

  where the sum is taken over all securities in the kth portfolio at time t

- Let $ME_{i,t}$ denote market value of security i at time t, then value-weighted returns are defined as (we measure $ME_{i,t}$ at portfolio formation)

$$r_{k,t} = \frac{\sum_{i=1}^{N_{k,t}} ME_{i,t-1} \times r_{i,t}}{\sum_{i=1}^{N_{k,t}} ME_{i,t-1}} \tag{5}$$

- Value-weighting is most appropriate when the securities are stocks, e.g., U.S. individual stocks from the CRSP sample
- Value-weighting alleviates issues with assigning *too large* weights to small and illiuid securities that are hard and expensive to trade

Step 4: Examining portfolio returns

- The main objective here is to determine whether there is a reliable cross-sectional relation between the sorting variable $F_{i,t}$ and future asset returns in the cross section

    1. The first step is to compute descriptive statistics for the portfolio (excess) returns and the long-short portfolio

Step 4: Examining portfolio returns, cont

2. We then look for monotonic relationships in the average returns between the first and the last portfolios

   $\rightarrow$ Patton and Timmermann (2010) provide a test to test for monotonicity in portfolio returns/factor exposure!

The Patton Timmermann test for monotonicity

- Let $\Delta_i = \mu_i - \mu_{i-j}$ denote the difference between the average return of portfolio i and $i - 1$
- We are then interested in testing whether all Δ_i's is greater than 0 or not. Meaning that

$$H_0 : \Delta \leq 0 \tag{6}$$
$$H_1 : \Delta > 0 \tag{7}$$

- The test statistic is simply $J_T = \min \Delta$
- The p-value is estimated using a block stationary bootstrap:
  1. Generate B random time-series of length T with average block length K
  2. For each random TS, calculate the average return, $\hat{\Delta}$ under the null-hypothesis ($\hat{\Delta} - \Delta$)
  3. J_T^b for a given bootstrap is then $J_T^b = \min(\hat{\Delta} - \Delta)$
  4. P-value is given as $\hat{p} = \frac{1}{B} \sum_{b=1}^{B} 1_{J_T^b > J_T}$

Step 4: Examining portfolio returns, cont

3. Next, we test whether the return pattern survives controlling for known risk factors identified in the asset pricing literature

- This amounts to testing intercepts (α) in time series regression using (some) asset pricing models

  $\rightarrow$ In the end the evaluation depends on the choice of asset pricing model! (FF3, FF5, Carhart 4-factor, etc.)

- We will illustrate the method and the impact of the choices using the momentum anomaly later in these slides

Bivariate Portfolio Sort

Bivariate portfolio sorts

- We now turn to bivariate (or double) portfolio sorts in which the universe of assets is sorted into portfolios based on two sorting variables rather than one
- Bivariate portfolio sorts are useful when we want to condition on (or control for) more than one sorting variable
- Bivariate sorts differ mainly in the construction of breakpoints and portfolio formation. The remaining steps are identical to the univariate case

Type of bivariate sorts

Types of bivariate sorts

In general, when considering bivariate portfolio sorts, we need to distinguish between independent and dependent sorts

- **Independent double sorts:** The ordering of the sorting variables does not matter
- **Dependent double sorts:** The ordering of the sorting variables is critically important

Independent double sort

- The independent double sort builds portfolios by sorting on two variables $F^1_{i,t}$ and $F^2_{i,t}$ independently
- We create n_{p_1} groups based on $F^1_{i,t}$ and n_{p_2} groups based on $F^2_{i,t}$ for a total of $n_{p_1} \times n_{p_2}$ portfolios
- The breakpoints for the two sorting variables are then defined as

$$\mathcal{B}^1_{k,t} = \text{Percentile}_{p_k}\left(F^1_{i,t}\right) \tag{8}$$

$$\mathcal{B}^2_{j,t} = \text{Percentile}_{p_j}\left(F^2_{i,t}\right), \tag{9}$$

- We are still facing the same choices for breakpoint determination as above:
  - Which assets should we use?
  - How many groups should we employ?
  - And what percentiles should be considered?

Building portfolios in the independent sort

- We create a total of $n_{p_1} \times n_{p_2}$ portfolios based on the groups identified using the sorting variables independently

- The portfolios are defined as the intersection of the groups based on the two sorting variables

$$P_{k,j,t} = \left\{ i \middle| \mathcal{B}^1_{k-1,t} \leq F^1_{i,t} \leq \mathcal{B}^1_{k,t} \right\} \cap \left\{ i \middle| \mathcal{B}^2_{j-1,t} \leq F^2_{i,t} \leq \mathcal{B}^2_{j,t} \right\} \tag{10}$$

  where $\cap$ is the intersection operator and $k = 1,2,\ldots,n_{p_1}$ and $j = 1,2,\ldots,n_{p_2}$ refers to the groups

- Portfolios are formed on the basis of the intersection of the groups of assets from each of the two independent sorts

- The remaining steps involving the calculation and examination of portfolio returns are equivalent to the univariate case. In a nutshell, we wish to establish if a reliable cross-sectional relationship exists

Dependent double sort

- The dependent double sort similarly builds portfolios by sorting on two variables $F_{i,t}^1$ and $F_{i,t}^2$. However, $F_{i,t}^1$ is now a control variable
- The main difference is that breakpoints for the second sorting variable in the dependent sort are formed *within* each group of the first sorting variable
- The n_{p_1} groups and breakpoints $\mathcal{B}_{k,t}^1$ for $k = 1,2,\ldots,n_{p_1} - 1$ for the first sorting variable is constructed identically to the independent sort case
- The breakpoints for the second sorting variable $F_{i,t}^2$ are now different and instead defined as

$$\mathcal{B}_{k,j,t}^2 = \text{Percentile}_{p_j}\left(F_{i,t}^2 \middle| \mathcal{B}_{k-1,t}^1 \leq F_{i,t}^1 \leq \mathcal{B}_{k,t}^1\right), \tag{11}$$

- Note that the order of the sorting variables is now critically important and a switch in ordering can lead to vastly different results

Building portfolios in the dependent sort

- All assets in the sample are first sorted into groups based on the breakpoints determined based on the first sorting variable $F^1_{i,t}$
- Assets in each of those groups are then sorted into portfolios based on the conditional breakpoints of the second sorting variable $F^2_{i,t}$
- The portfolios are defined as the intersection of the conditional groups based on the two sorting variables

$$P_{k,j,t} = \left\{ i | \mathcal{B}^1_{k-1,t} \leq F^1_{i,t} \leq \mathcal{B}^1_{k,t} \right\} \cap \left\{ i | \mathcal{B}^2_{k,j-1,t} \leq F^2_{i,t} \leq \mathcal{B}^2_{k,j,t} \right\} \qquad (12)$$

where $\cap$ is the intersection operator and $k = 1,2,\ldots,n_{p_1}$ and $j = 1,2,\ldots,n_{p_2}$ refers to the groups

- The remaining steps involving the calculation and examination of portfolio returns are equivalent to the univariate case

The CRSP universe

The CRSP stock file

- We want to study cross-sectional patterns in expected returns using an appropriate universe of financial assets
- The main source for data on U.S. individual stocks is the Center for Research in Security Prices (CRSP) stock file
  - CRSP is maintained by the University of Chicago's Booth School of Business
  - The CRSP tape provides data from December 31, 1925 up to today
  - The CRSP tape is hosted through Wharton Research Data Services (WRDS)
  - Also contains data on market indices, stock market factors, and bonds
  - The access link is here: https://wrds-www.wharton.upenn.edu
- For more information and details about log-in (signing up) and how to use the CRSP web-based access at WRDS, see crspNotes.pdf on Brightspace

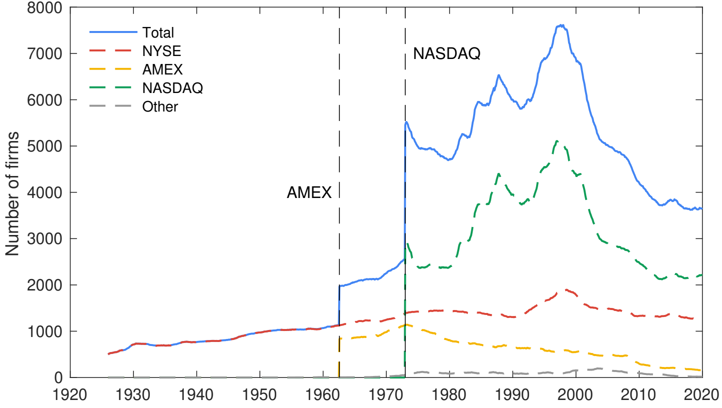
Overview of data

- CRSP contains monthly and daily data on U.S. individual stocks. The most important stock variables for our purpose here are

  1. **PERMNO:** Every stock issue is assigned a unique PERMNO that does not change over time, even if the company name, ticker, or exchange do. PERMNO is the principal identifier of a stock in CRSP
  2. **SHRCD:** Every stock is issued a share code and we can use it to identify U.S. common stocks (SHRCD 10 and 11)
  3. **EXCHCD:** A stock's exchange code indicates the exchange on which the security is listed, e.g., NYSE, AMEX, or NASDAQ (EXCHCD 1, 2, and 3, respectively)
  4. **PRC:** The price of the security is the closing price or the negative bid/ask average for a trading day (so always use the absolute price)
  5. **RET:** The holding-period return of a security including dividend payments and share repurchases and takes splits into account
  6. **SHROUT:** The number of shares outstanding is the number of publicly held shares (recorded in thousands) and is useful for computing market equity and value-weighted portfolio returns

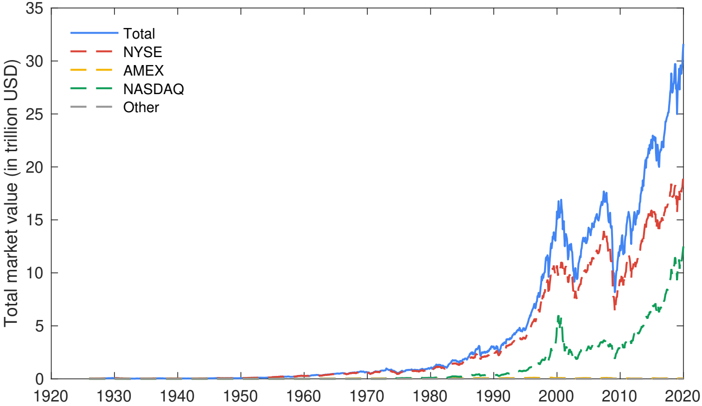
Number of firms in CRSP by exchange

- The number of firms available in the CRSP sample varies considerably over time and across exchanges
- AMEX is included July 1962 and NASDAQ in December 1972



Market equity of firms in CRSP by exchange

- The total market value of stocks listed across the different exchanges differ greatly. Most originates from stocks on NYSE and NASDAQ, whereas very little originates from AMEX and other stocks



Empirical illustrations: Momentum

Momentum portfolios

- To illustrate univariate portfolio sorting, we consider the construction of the momentum anomaly (Jegadeesh and Titman, 1993) using the CRSP sample
- We consider common stocks (SHRCD 10 and 11) listed on the NYSE, AMEX, and NASDAQ exchanges (EXCHCD 1, 2, and 3) from January 1986 to December 2019

Momentum signal

We define the momentum signal as in Carhart (1997) and Asness et al. (2013a), where momentum at time $t-1$ is defined as the cumulative return from $t-12$ to $t-2$

$$F_{i,t-1} = \prod_{h=0}^{10} (1 + r_{i,t-12+h}) \tag{13}$$

where $r_{i,t}$ is the return on stock i at time t, and we skip the most recent month to avoid short-term reversal effects (Jegadeesh, 1990, Lo and MacKinlay, 1990)

Choices and requirements for stocks

- We illustrate the impact of the choices open to the researcher by building momentum portfolios using

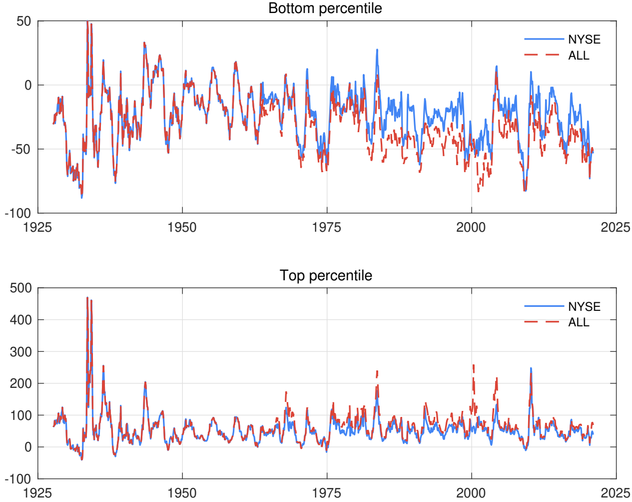
  1. Breakpoints based on NYSE and ALL stocks
  2. Equal- and value-weighted portfolio returns

- In our implementations, we follow Kenneth R. French and require the following for a stock to be included in the sample

  1. Portfolios are re-balanced every month
  2. The price at time $t-13$ is not missing
  3. The return at time $t-2$ is not missing
  4. Market equity data at time $t-1$ is not missing

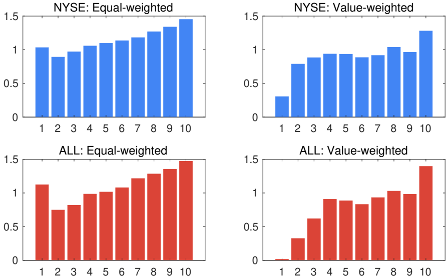
Breakpoints

- First, consider the differences in top and bottom percentiles when using NYSE and ALL stocks, respectively



Momentum returns

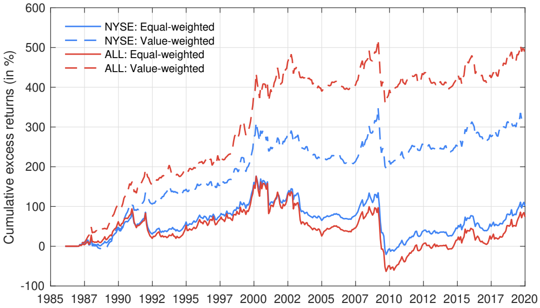
- Consider next the differences in return patterns originating from breakpoint choices and using equal- and value-weighted portfolio returns



- Applying the Patton and Timmermann (2010) test yields that we cannot reject except for VW-all

Cumulative excess momentum returns

- Consider also the cumulative excess returns to a momentum strategy in stocks for the same breakpoint choices and equal- and value-weighted portfolio excess returns



Momentum returns

- Last, we consider descriptive statistics for the NYSE-based, value-weighted momentum portfolio excess returns and risk-adjusted excess returns using the Fama and French (1993) three-factor model

$$r_{k,t} - r_{f,t} = \alpha_k + b_{MKT} MKT_t + b_{SMB} SMB_t + b_{HML} HML_t + \varepsilon_{k,t} \qquad (14)$$

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	MOM
					Panel A: Descriptive statistics						
Mean	0.62	6.41	**7.55**	**8.21**	**8.18**	**7.58**	**7.95**	**9.41**	**8.53**	**12.31**	**11.69**
	[0.11]	[1.55]	**[2.26]**	**[2.82]**	**[3.14]**	**[2.96]**	**[3.54]**	**[3.91]**	**[3.19]**	**[3.40]**	**[2.38]**
Std	30.52	22.39	18.87	16.49	15.12	14.69	14.31	14.09	15.33	20.58	26.57
Skew	0.59	0.15	0.17	-0.45	-0.56	-0.93	-0.97	-0.74	-0.98	-0.62	-1.44
Kurt	6.96	6.77	7.06	5.44	6.22	7.12	7.26	5.55	7.28	5.47	10.51
SR	0.02	0.29	0.40	0.50	0.54	0.52	0.56	0.67	0.56	0.60	0.44
					Panel B: Risk-adjusted returns						
α	**-12.53**	**-4.31**	-1.88	-0.42	0.22	-0.27	0.64	**2.26**	1.15	**4.51**	**17.04**
	[-4.11]	**[-2.02]**	[-1.05]	[-0.35]	[0.20]	[-0.27]	[0.56]	**[2.71]**	[1.01]	**[2.63]**	**[4.26]**
MKT	**1.56**	1.27	1.10	1.02	0.96	0.94	0.90	0.89	0.94	**1.04**	**-0.52**
	[12.48]	[15.95]	[21.57]	[30.09]	[29.48]	[34.32]	[24.43]	[29.28]	[20.55]	**[20.63]**	**[-3.29]**
SMB	**0.41**	0.11	-0.02	-0.09	-0.10	-0.09	-0.16	-0.09	-0.11	**0.38**	-0.03
	[2.55]	[0.83]	[-0.15]	[-1.55]	[-1.48]	[-2.02]	[-3.48]	[-2.35]	[-1.67]	**[5.69]**	[-0.14]
HML	**0.34**	0.37	0.40	0.33	0.25	0.27	0.16	0.13	0.03	**-0.30**	**-0.64**
	[1.46]	[2.61]	[4.07]	[4.48]	[5.05]	[5.49]	[3.53]	[2.17]	[0.59]	**[-3.13]**	**[-2.10]**
Adj. R^2	62.92	70.96	73.56	81.61	84.96	86.58	83.17	83.96	81.08	76.73	11.06

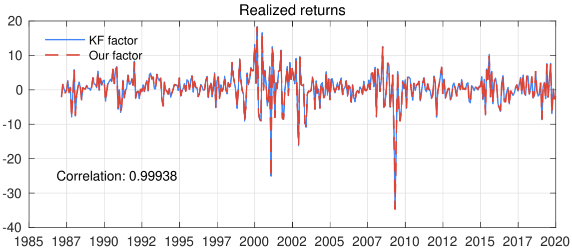
A momentum factor

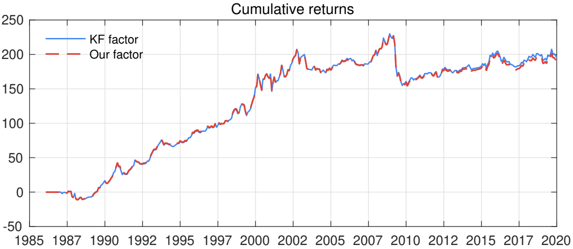
- We illustrate an independent double sort by replicating the momentum factor (Carhart, 1997) available on Kenneth French's data library
- The full universe of stocks consists of common shares (SHRCD 10 and 11) listed on NYSE, AMEX, and NASDAQ
- We independently sort stocks based on size (ME) and momentum (MOM), where we consider two size groups determined using the median NYSE size as the breakpoint and three momentum groups using breakpoints determined from the 30th and 70th NYSE percentiles
- The intersections provide us with six portfolios: Big Losers (BL), Big Neutrals (BN), Big Winners (BW), Small Losers (SL), Small Neutrals (SN), and Small Winners (SW)
- The momentum factor is then constructed from the six portfolios as follows

$$MOM = \frac{1}{2}\left[SW + BW\right] - \frac{1}{2}\left[SL + BL\right], \tag{15}$$

Comparison with Kenneth French

- The momentum factor constructed here as a correlation of 0.9993 with the factor obtained from Kenneth French, and the series are very similar





Empirical illustrations: Macroeconomic Uncertainty

The relevance for you

- The paper we will go now through is a (good) example of how you can construct a factor from macro information
- You can apply the same method and type of discussion with any macro variable of interest
- But always ask yourself; why does it make intuitive sense that the variable is priced in financial markets?

  $\rightarrow$ For now, we will consider the case of the macroeconomic uncertainty index of Jurado et al. (2015)

Motivation

- The ICAPM model of Merton (1973) suggests that investors seek to hedge changes in investment and consumption opportunity sets
- An implication is any variable that correlates with these opportunity sets should be priced in the cross-section
- Prior studies find a link economic uncertainty and the real economy in addition to asset prices (Bloom, 2009, Drechsler, 2013, Augustin and Tédongap, 2021, Ludvigson et al., 2021)

  $\rightarrow$ MU should correlate with the opportunity sets

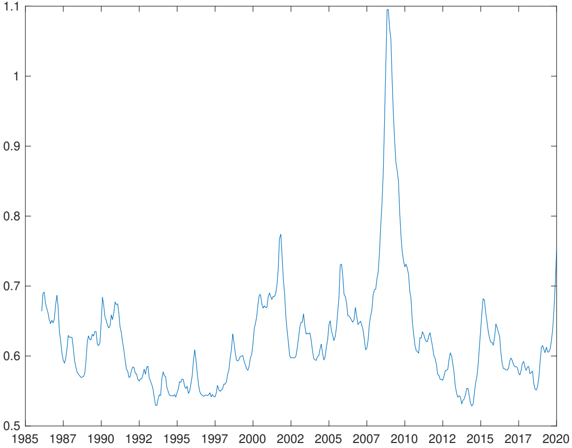
The macroeconomic uncertainty index

- Bali et al. (2017) examine the macroeconomic uncertainty measure of Jurado et al. (2015)
- Jurado et al. (2015) defines uncertainty as volatility on forecast errors, i.e.,

$$\mathcal{U}_{j,t}^{h} = \sqrt{E\left[\left(y_{t+h,j} - E_t(y_{t+h,j})\right)^2 \mid I_t\right]} \tag{16}$$

- The measure is then aggregated (equal-weighted) across the $j \in J$ variables
- Note, that the index is not a vintage dataset meaning that we can discuss whether the index is in investors' information set

The macrouncertainty index



The MU exposure factor

- To examine whether MU is priced in the cross-section of US stocks, Bali et al. (2017) estimates the following pricing model for each asset i

$$\begin{aligned} R^e_{i,t} &= \alpha_i + \beta_{MU,i} MU_t + \beta_{MKT,i} MKT_t + \beta_{SMB,i} SMB_t \\ &+ \beta_{HML,i} HML_t + \beta_{UMD,i} UMD_t + \beta_{LIQ,i} LIQ_t \\ &+ \beta_{IA,i} R_{IA,t} + \beta_{ROE,i} R_{ROE,t} + \varepsilon_{i,t} \end{aligned} \tag{17}$$

- Using a rolling window of 5 year (60 months for 9 parameters)
- The β_{MU} estimates are then saved for the portfolio sort

data

- We consider the FF5 model for MKT, SMB, HML, AI, ROE factors
- The UMB from Carhart (1997) obtained from Kenneth French
- The liquidity factor from Pástor and Stambaugh (2003)
- We consider, again, the CRSP dataset from 1986 to 2019

Univariate portfolio analysis

	1	2	3	4	5	6	7	8	9	10	L-S
						VW					
Excess return	1.12	1.01	0.76	0.77	0.88	0.80	0.58	0.67	0.61	0.75	-0.37
	(2.75)	(3.19)	(2.69)	(3.39)	(4.06)	(3.94)	(2.62)	(2.91)	(2.20)	(1.88)	(-1.20)
α_{FF5}	0.35	0.14	-0.07	-0.08	0.13	0.03	-0.15	-0.11	-0.07	-0.02	-0.37
	(1.51)	(0.66)	(-0.58)	(-1.04)	(1.43)	(0.36)	(-2.60)	(-1.28)	(-0.62)	(-0.13)	(-1.11)
						EW					
Excess return	1.56	1.31	1.25	1.06	1.06	1.09	0.98	0.96	1.02	1.04	-0.52
	(3.33)	(3.86)	(4.24)	(3.92)	(4.07)	(4.39)	(3.75)	(3.56)	(3.16)	(2.50)	(-2.15)
α_{FF5}	0.75	0.39	0.37	0.17	0.19	0.24	0.15	0.12	0.16	0.28	-0.47
	(2.60)	(2.68)	(3.32)	(1.90)	(2.41)	(3.25)	(1.83)	(1.37)	(1.50)	(1.72)	(-1.94)

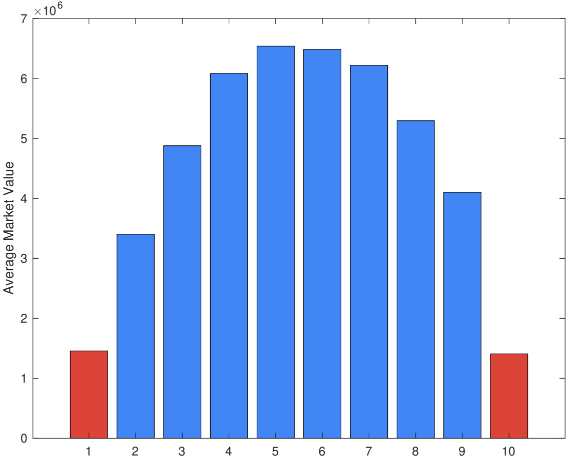
Discussion of choices

- Bali et al. (2017) make some choices in the construction of the analysis:
  1. The model in eq. (17)
  2. The 5-year rolling window
  3. They calculate breakpoints based on the entire cross-section instead of NYSE(Hou et al., 2020)
  4. They measure economic activity by a three month moving average of CFNAI

Different model specification

	1	2	3	4	5	6	7	8	9	10	L-S
						EW					
Excess returns	1.51	1.26	1.15	1.04	1.09	1.07	0.98	0.98	1.08	1.16	-0.34
	3.27	3.79	3.92	3.94	4.33	4.31	3.72	3.43	3.38	2.70	-1.33
α_{FF5}	0.67	0.36	0.29	0.19	0.27	0.22	0.13	0.09	0.22	0.39	-0.28
	2.11	2.42	2.40	2.02	3.36	3.06	1.60	1.22	2.37	2.39	-0.97

Market value across portfolios



NYSE based breakpoints

	1	2	3	4	5	6	7	8	9	10	L-S
						VW					
Excess return	0.96 (2.58)	0.95 (3.43)	0.79 (3.21)	0.77 (3.36)	0.90 (4.32)	0.82 (3.89)	0.56 (2.55)	0.59 (2.56)	0.71 (2.86)	0.75 (2.30)	-0.20 (-0.79)
α_{FF5}	0.14 (0.65)	0.10 (0.81)	-0.09 (-0.89)	-0.08 (-0.87)	0.20 (2.10)	0.02 (0.27)	-0.16 (-2.02)	-0.17 (-1.88)	-0.02 (-0.24)	0.03 (0.21)	-0.12 (-0.39)

Empirical illustrations: Carry

Currency carry trade

- Lustig et al. (2011a) and Menkhoff et al. (2012) show that investing (borrowing) in high (low) interest rate countries provide a large excess return
- Studying the currency carry trade requires a broad cross-section of currencies and we will make use of the data from Verdelhan (2018a) for illustration
- Let s_t^k and f_t^k denote the log spot and one-month forward exchange rate, respectively, then log currency excess returns are defined as

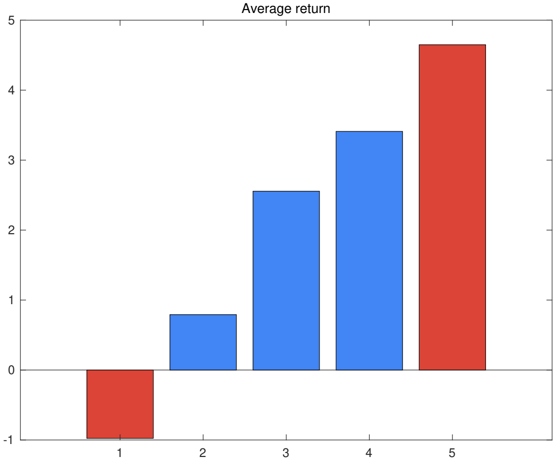
$$rx_{t+1}^k = \underbrace{f_t^k - s_t^k}_{\text{Forward discount}} - \underbrace{\Delta s_{t+1}^k}_{\text{Spot exchange rate change}} \tag{18}$$

$$\approx \underbrace{i_t^k - i_t}_{\text{Interest differential}} - \underbrace{\Delta s_{t+1}^k}_{\text{Spot exchange rate change}} \tag{19}$$

where i_t^k and i_t denote the foreign and domestic interest rates, respectively

The Carry Strategy

- The Carry strategy delivers the following average annualized returns



The Carry Strategy

	1	2	3	4	5	HML
Mean	-0.98	0.79	2.56	3.41	4.65	5.63
	(-0.69)	(0.63)	(1.70)	(2.28)	(2.54)	(3.74)
Std	7.77	7.07	7.82	8.09	9.72	8.64
SR	-0.13	0.11	0.33	0.42	0.48	0.65

- The HML is highly significant!

Volatility innovations and carry trades

Menkhoff et. al. (2011)

- We will now go through the article of Menkhoff et al. (2012a) which provides a nice example of how you can test for a specific risk factor is priced in the cross-section
- The article, furthermore, provides an example of to test whether a non-traded factor is priced in the cross-section of some asset (in this case currencies)
- Said simple: the authors examine whether global exchange volatility can explain the cross-section of portfolios sorted on interest rates

Why consider Volatility?

- The relation between risk and return is at the core of empirical asset pricing
- Time-varying market volatility induced changes in the investment opportunity set by changing expected returns and/or the risk-return trade-off (again think of the ICAPM model)
- Ang et al. (2006) documents that stocks with high sensitivities to innovations in aggregate volatility have low average returns - consistent with risk-based asset pricing
- Menkhoff et al. (2012a) essentially copy-paste the analysis of Ang et al. (2006) into the currency framework
- The findings of the paper is similar: high-interest rate currencies have a negative relation to aggregated FX vol and, hence, deliver a negative return due to increases in volatility
- The profitability of the Carry trade is, hence, compensation for risk

Initial comments for the analysis

- To ease the analysis (and avoid confusion), we will ignore transaction costs
- We will, furthermore, consider discrete returns instead of $\log$ returns (and do the analysis without approximations)
- We follow Lustig et al. (2011b) and exclude currencies in which the CIP has been documented to be violated
- We will consider the same Carry returns as from before

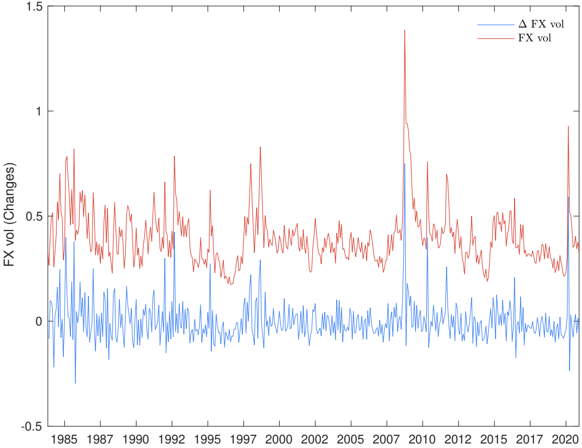
Global Foreign Exchange Volatility

- To proxy the volaility Menkhoff et al. (2012a)

$$\sigma_t^{FX} = \frac{1}{T_t} \sum_{\tau \in T_t} \left[\sum_{k \in K_\tau} \frac{\mid r_\tau^k \mid}{K_\tau} \right] \tag{20}$$

- They consider absolute returns instead of realized volatility to minimize the impact of outliers
- Menkhoff et al. (2012a) focus on volatility innovations which they define as residuals from an AR(1) process

Global Foreign Exchange Volatility



Motivational evidence for volatility

- Let's first examine a motivating plot on the relationship between carry returns and vol innovations:

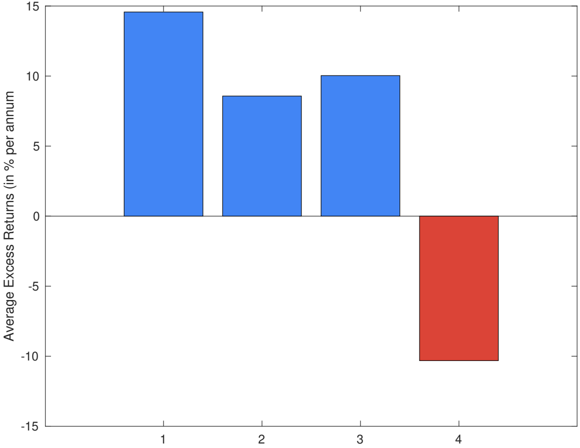


Fig. 1: Carry performace across Δ vol quartiles

A linear SDF

- To test for whether vol innovations can explain the variation in Carry portfolios, Menkhoff et al. (2012a) consider the following linear SDF

$$M_t = (1 - b_{Dol}(DOL_t - \mu_{Dol}) - b_{VOL}\Delta\sigma_{FX,t}) \tag{21}$$

- where the DOL factor is now defined as the cross-sectional average of Carry portfolios
- To estimate the parameters, they apply GMM

Moment conditions

- Denote $z_t = (rx_t, h_t)$ where $h_t = (DOL_t, \sigma_{FX,t})$
- More specifically, they consider the following moment conditions

$$g(z_t, \theta) = \begin{bmatrix} [1 - b'(h_t - \mu)]\, rx_t \\ h_t - \mu \\ vec((h_t - \mu)(h_t - \mu)') - vec(\Sigma_h) \end{bmatrix} \tag{22}$$

- As weighting matrix, they consider the identity matrix for the SDF and a large weight (I have set it to 1000) for the additional
- They consider a HAC estimator for the long-run covariance matrix with Andrews (1991) lag length

From SDF loading to risk premium

- Remember that the risk premium λ is given as

$$\lambda = \Sigma_h b \tag{23}$$

- We can then apply the delta method to conduct inference on the risk premium estimates
- Denote

$$\theta = \begin{bmatrix} vec(b) \\ vec(\Sigma_h) \end{bmatrix} \tag{24}$$

- The delta method state that if

$$\hat{\theta} \sim^d \mathcal{N}(\theta, \Sigma_\theta) \tag{25}$$

then

$$g(\hat{\theta}) \sim^d \mathcal{N}(g(\theta), g'(\theta)\Sigma_\theta(g'(\theta))^\intercal) \tag{26}$$

In our case

- In our case

$$g'(\theta) = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} & b_{1,1} & b_{2,2} & 0 & 0 \\ \Sigma_{2,1} & \Sigma_{2,2} & 0 & 0 & b_{1,1} & b_{2,2} \end{bmatrix} \tag{27}$$

- and Σ_θ is the covariance matrix of our estimates

Asset pricing test

- We are now ready to test whether vol innovations can explain cross-sectional variation in Carry portfolios. The table below provides the results

	DOL	VOL	J-test
b	-0.01	-4.71	1.38
	(-0.18)	(-2.31)	[0.71]
λ	0.15	-0.05	
	(1.60)	(-3.37)	

Asset pricing test, 2.0

- We can also examine whether the specified SDF delivers significant pricing errors (note a slightly different focus in Menkhoff et al. (2012a)):

Portfolio	α	β_{DOL}	β_{σ}	R^2
1	-0.25	0.98	4.24	76.90
	(-4.84)	(18.29)	(7.59)	
2	-0.09	0.89	1.01	77.72
	(-1.79)	(19.63)	(2.04)	
3	0.04	1.01	-0.01	84.89
	(0.87)	(24.13)	(-0.02)	
4	0.10	1.04	-0.81	83.93
	(1.98)	(24.11)	(-1.52)	
5	0.20	1.09	-4.43	70.68
	(2.51)	(19.79)	(-4.28)	

Factor mimicking portfolio

- To better measure ex post exposure to aggregate FX volatility risk at a monthly frequency, Menkhoff et al. (2012a) follow Breeden et al. (1989) and Ang et al. (2006) and build a factor-mimicking portfolio from the regression

$$\Delta\sigma_t^{FX} = c + b' rx_t + u_t \tag{28}$$

- So we construct a portfolio of the carry portfolios that maximize the correlation between the factor of interest and mimicking portfolio
- The factor-mimicking portfolio is then constructed as the fitted values (less the intercept c) from the regression in (28), i.e.,

$$rx_t^{FM} = \hat{b}' rx_t \tag{29}$$

- One could essentially use any set of assets with sufficient return dispersion to construct the mimicking portfolio

Factor mimicking portfolio

- In our case, the FM weights are given as

$$rx_t^{FM} = 0.25rx_t^1 - 0.07rx_t^2 - 0.08rx_t^3 - 0.05rx_t^1 - 0.13rx_t^5 \tag{30}$$

- The mimicking portfolio loads most heavily on the extreme deciles and is somewhat decreasing from portfolio 2 through 5
- This implies that portfolio 1 provides a hedge against FX vol increases

Asset pricing tests 2.0 with the factor mimicking portfolio

- We can then examine whether the factor mimicking portfolio is priced

	DOL	RX^{FM}	J-test
b	0.01	-0.39	1.38
	(-0.08)	(-1.07)	[0.71]
λ	0.46	-0.10	
	(3.44)	(-3.10)	

Vol exposure and the cross-section of currencies

- We can also construct a portfolio sort based on vol innovation exposures
- Menkhoff et al. (2012a) consider a 36 rolling window of the following regression to measure the exposure

$$rx_{t,j} = \alpha + \beta_{DOL} DOL_t + \beta_{VOL} \Delta\sigma_t^{FX} + \eta_{t,j} \tag{31}$$

- Note, that they only rebalance every 6th month (nonstandard choice)
- Instead, we will follow the standard approach and rebalance every month

Vol exposure and the cross-section of currencies

	1	2	3	4	5	L-H
Mean	2.00	2.03	2.03	1.24	0.79	1.21
	(1.12)	(1.32)	(1.70)	(0.89)	(0.57)	(0.78)
STD	9.74	8.28	7.01	6.97	7.37	9.19
SR	0.21	0.24	0.29	0.18	0.11	0.13

- Not that convincing evidence!

Potential projects

Stocks

In addition to the papers cited already, you can look into:

- Default risk (Vassalou and Xing, 2004)
- volatility (Ang et al., 2006)
- Liquidity (Pástor and Stambaugh, 2003)
- Financial constraints (Owen et al., 2001)
- Tail risk (Kelly and Jiang, 2014, Bali et al., 2014)
- Profitablity (Fama and French, 2015)
- Bid-ask spreads (Corwin and Schultz, 2012)
- Financial Intermediation (Adrian et al., 2014)
- ETC....

Hou et al. (2020) examine 452 anomalies so you have an endless list of oppotunities!

Corporate bonds

- Bond maturity (Baker et al., 2003)
- MU (Bai et al., 2021)
- Long-run reversals (Bali et al., 2021)
- VaR (Bai et al., 2019)

Commodities

- Carry/basis (Yang, 2013, Koijen et al., 2018)
- Value and momentum (Asness et al., 2013b)
- Low beta (Frazzini and Pedersen, 2014)
- Inventories (Gorton et al., 2013)

Factors

- You can naturally also try to sort anomalies based on some chacteristics, e.g,. factor momentum (Ehsani and Linnainmaa, 2019)
- Other examples; Value spread between L-S leg, momentum gab, etc.,

Currencies

- Macro-economic difference: output gap (Colacito et al., 2020), momentum in economic fundamentals (Dahlquist and Hasseltoft, 2020), global imbalances (Corte et al., 2016)
- Traditional: Momentum(Menkhoff et al., 2012b), value (Menkhoff et al., 2017)
- Financial variables: currency volatility risk premia (Della Corte et al., 2016), Order flows (Menkhoff et al., 2016), Dollar beta Verdelhan (2018b)

➜ In other words: see the website of Lucio Sarno...

Potential projects

- Portfolio sorting is not unique to stocks, but the method is equally applicable to currencies, bonds, commodities, and pretty much any asset with a return
- Sort a universe of assets into portfolios based on a *new* sorting variable and evaluate whether the inherent risk is priced in the cross-section
- Re-evaluate an existing sorting factor using updated data, more subsamples, and/or different choices for the data and construction of portfolios
- Construct an existing set of anomaly portfolios and investigate a well-motivated risk-based (or behavioral) explanation
- Consider new double sorts where you take one well known variable together with a new variable (e.g., attention, sentiment, etc,) and explore their joint dynamics (see, e.g., Medhat and Schmeling (2020) for a recent example)
- Build a new cross-section of test assets to evaluate existing risk factors and asset pricing models

References

Adrian, T., E. Etula, and T. Muir (2014). Financial intermediaries and the cross-section of asset returns. *The Journal of Finance 69*(6), 2557–2596.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, 817–858.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *The journal of finance 61*(1), 259–299.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen (2013a). Value and momentum everywhere. *Journal of Finance 68*(3), 929–985.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen (2013b). Value and momentum everywhere. *The Journal of Finance 68*(3), 929–985.

Augustin, P. and R. Tédongap (2021). Disappointment aversion, term structure, and predictability puzzles in bond markets. *Management Science 67*(10), 6266–6293.

Bai, J., T. G. Bali, and Q. Wen (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics 131*(3), 619–642.

Bai, J., T. G. Bali, and Q. Wen (2021). Is there a risk-return tradeoff in the corporate bond market? time-series and cross-sectional evidence. *Journal of Financial Economics 142*(3), 1017–1037.

Baker, M., R. Greenwood, and J. Wurgler (2003). The maturity of debt issues and predictable variation in bond returns. *Journal of Financial Economics 70*(2), 261–291.

Bali, T. G., S. J. Brown, and Y. Tang (2017). Is economic uncertainty priced in the cross-section of stock returns? *Journal of Financial Economics 126*(3), 471–489.

Bali, T. G., N. Cakici, and R. F. Whitelaw (2014). Hybrid tail risk and expected stock returns: when does the tail wag the dog? *The Review of Asset Pricing Studies 4*(2), 206–246.

Bali, T. G., A. Subrahmanyam, and Q. Wen (2021). Long-term reversals in the corporate bond market. *Journal of Financial Economics 139*(2), 656–677.

Bloom, N. (2009). The impact of uncertainty shocks. *econometrica 77*(3), 623–685.

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger (1989). Empirical tests of the consumption-oriented capm. *The Journal of Finance 44*(2), 231–262.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

Colacito, R., S. J. Riddiough, and L. Sarno (2020). Business cycles and currency returns. *Journal of Financial Economics 137*(3), 659–678.

Corte, P. D., S. J. Riddiough, and L. Sarno (2016). Currency premia and global imbalances. *The Review of Financial Studies 29*(8), 2161–2193.

Corwin, S. A. and P. Schultz (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *The Journal of Finance 67*(2), 719–760.

Dahlquist, M. and H. Hasseltoft (2020). Economic momentum and currency returns. *Journal of Financial Economics 136*(1), 152–167.

Della Corte, P., T. Ramadorai, and L. Sarno (2016). Volatility risk premia and exchange rate predictability. *Journal of Financial Economics 120*(1), 21–40.

Drechsler, I. (2013). Uncertainty, time-varying fear, and asset prices. *The Journal of Finance 68*(5), 1843–1889.

Ehsani, S. and J. T. Linnainmaa (2019). Factor momentum and the momentum factor. Technical report, National Bureau of Economic Research.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks 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), 1–22.

References

Frazzini, A. and L. H. Pedersen (2014). Betting against beta. *Journal of Financial Economics 111*(1), 1–25.

Gorton, G. B., F. Hayashi, and K. G. Rouwenhorst (2013). The fundamentals of commodity futures returns. *Review of Finance 17*(1), 35–105.

Hou, K., C. Xue, and L. Zhang (2020). Replicating anomalies. *The Review of Financial Studies 33*(5), 2019–2133.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance 45*(3), 881–898.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *American Economic Review 105*(3), 1177–1216.

Kelly, B. and H. Jiang (2014). Tail risk and asset prices. *The Review of Financial Studies 27*(10), 2841–2871.

Koijen, R. S., T. J. Moskowitz, L. H. Pedersen, and E. B. Vrugt (2018). Carry. *Journal of Financial Economics 127*(2), 197–225.

Lo, A. W. and C. MacKinlay (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies 3*(2), 175–205.

Ludvigson, S. C., S. Ma, and S. Ng (2021). Uncertainty and business cycles: exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics 13*(4), 369–410.

Lustig, H., N. Roussanov, and A. Verdelhan (2011a). Common risk factors in currency markets. *Review of Financial Studies 24*(11), 3731–37777.

Lustig, H., N. Roussanov, and A. Verdelhan (2011b). Common risk factors in currency markets. *The Review of Financial Studies 24*(11), 3731–3777.

Medhat, M. and M. Schmeling (2020). Short-term momentum. *Review of Financial Studies*, Forthcoming.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012a). Carry trades and global foreign exchange volatility. *The Journal of Finance 67*(2), 681–718.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012b). Currency momentum strategies. *Journal of Financial Economics 106*(3), 660–684.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2016). Information flows in foreign exchange markets: Dissecting customer currency trades. *The Journal of Finance 71*(2), 601–634.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2017). Currency value. *The Review of Financial Studies 30*(2), 416–441.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrumpf (2012). Carry trades and global foreign exchange volatility. *Journal of Finace 67*(2), 681–718.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 867–887.

Owen, L., C. Polk, and J. Saa-Requejo (2001). Financial constraints and stock returns. *Review of Financial Studies 14*, 529–554.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political economy 111*(3), 642–685.

Patton, A. J. and A. Timmermann (2010). Monotonicity in asset returns: New tests with applications to the term structure, the capm, and portfolio sorts. *Journal of Financial Economics 98*(3), 605–625.

Vassalou, M. and Y. Xing (2004). Default risk in equity returns. *The journal of finance 59*(2), 831–868.

Verdelhan, A. (2018a). The share of systematic variation in bilateral exchange rates. *Journal of Finance 73*(1), 375–418.

Verdelhan, A. (2018b). The share of systematic variation in bilateral exchange rates. *The Journal of Finance 73*(1), 375–418.

Yang, F. (2013). Investment shocks and the commodity basis spread. *Journal of Financial Economics 110*(1), 164–184.

Risk factors out-of-sample

Empirical Asset Pricing

Mads Markvart Kjær

Department of Economics and Business Economics, Aarhus University

E-mail: mads.markvart@econ.au.dk

Spring 2023

Risk factors out-of-sample

- There is great debate about the number and identification of reliable risk factors and cross-sectional return predictors (Harvey et al., 2016, Harvey, 2017, McLean and Pontiff, 2016, Linnainmaa and Roberts, 2018, Engelberg et al., 2018, Calluzzo et al., 2019, Chen and Zimmermann, 2020, Chen, 2021a, Chen and Zimmermann, 2021, Jensen et al., 2021, Harvey and Liu, 2021)

The opposing sides of the debate

Dark side: Factors are false

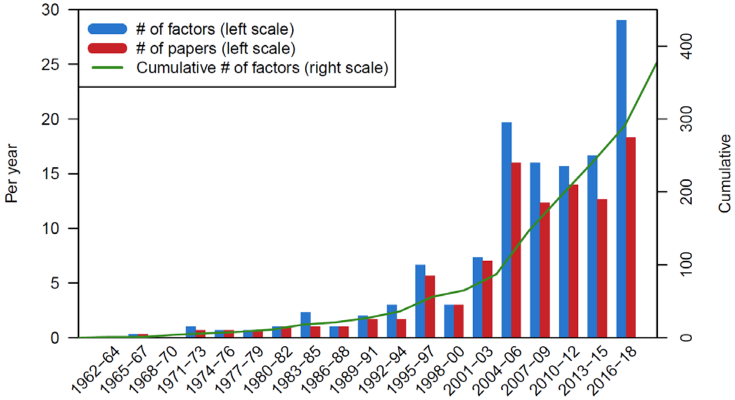
1. Data-mining
2. p-hacking
3. Chance result
4. Replication crisis

Bright side: Factors are real

1. Economically motivated
2. Robustness checks
3. Mispricing
4. Factors replicate

Factor production is rampant

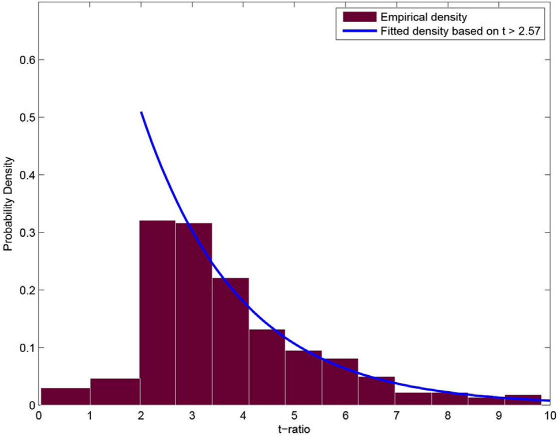
- Harvey and Liu (2019) provide a census of the proverbial factor zoo (Cochrane, 2011, Harvey et al., 2016) Google sheet with factors
- The below figure details the factor production from 1963–2018 and clearly indicates a strong increase (382 factors in top journals alone)



* Journals published through December 2018. Data collection in January 2019.

Distribution of reported t-statistics

- Harvey (2017) argues that the distribution of reported t-statistics is indicative of a publication bias and p-hacking
- In particular, note that t-statistics in the range of 2.0 to 2.57 is almost the same as the number reporting in the range of 2.57 to 3.14. Also, notice that very few papers with negative results (t-statistic less than 2.00) are published



McLean and Pontiff (2016)

McLean and Pontiff (2016)

- McLean and Pontiff (2016) ask the following question: Does cross-sectional return predictability typically persists post-publication?
- To investigate this, they consider 97 cross-sectional return predictors (or risk factors) over three distinct periods
  1. The original study's sample period
  2. The out-of-sample period (post-sample, pre-publication)
  3. The post-publication period
- Examining the behavior of returns over these three periods, we can distinguish between three possible explanations
  1. Statistical bias $\Rightarrow$ Predictability should disappear out-of-sample
  2. Rational risk $\Rightarrow$ Predictability should be the same in-sample, out-of-sample, and post-publication
  3. Mispricing $\Rightarrow$ If arbitrage is costless, then it should disappear completely. If arbitrage is costly, then the effect should at least decay post-publication

Summary of main findings

- The main findings in McLean and Pontiff (2016) can be summarized as follows

  1. The average predictor's long-short return declines by 26% out-of-sample
     - One can view this as an upper bound on the effect of statistical biases!
  2. The average predictor's long-short return shrinks 58% post-publication
     - Together with the 26% out-of-sample decline, this implies a lower bound on the publication effect of about 32%
     - Rejection of the hypothesis that return predictability disappears completely and that predictability post-publication is unchanged
  3. The decay in portfolio returns is larger for predictor portfolios with higher in-sample returns and t-statistics
  4. Post-publication returns are lower for predictors that are less costly to arbitrage, i.e., portfolios concentrated in liquid and low idiosyncratic risk stocks
  5. Publication affects the correlation between predictor portfolio returns. (Yet-to-be) published predictors are highly correlated with (yet-to-be) published

$\Rightarrow$ The empirical evidence suggests that investors learn about mispricing from academic publications

Summary statisics

- The below table provides an overview of the characteristics of the data. Note that only 85 of the 97 predictors have a t-statistic above 1.5

Number of predictor portfolios	97
Predictors portfolios with t-statistic > 1.5	85 (88%)
Mean publication year	2000
Median publication year	2001
Predictors from finance journals	68 (70%)
Predictors from accounting journals	27 (28%)
Predictors from economics journals	2 (2%)
Mean portfolio return in-sample	0.582
Standard deviation of mean in-sample portfolio return	0.395
Mean observations in-sample	323
Mean portfolio return out-of sample	0.402
Standard deviation of mean out-of-sample portfolio return	0.651
Mean observations out-of-sample	56
Mean portfolio return post-publication	0.264
Standard deviation of mean post-publication portfolio return	0.516
Mean observations post-publication	156

Cross-sectional predictors

- In their Internet Appendix, they provide a full list of predictors and their classification into four broad categories (plus definitions)

Event	Market	Valuation	Fundamental
Change in Asset Turnover	52-Week High	Advertising/MV	Accruals
Change in Profit Margin	Age-Momentum	Analyst Value	Age
Change in Recommendation	Amihud's Measure	Book-to-Market	Asset Growth
Chg. Forecast + Accrual	Beta	Cash Flow/MV	Asset Turnover
Debt Issuance	Bid/Ask Spread	Dividends	Cash Flow Variance
Dividend Initiation	Coskewness	Earnings-to-Price	Earnings Consistency
Dividend Omission	Idiosyncratic Risk	Enterprise Component of B/P	Forecast Dispersion
Dividends	Industry Momentum	Enterprise Multiple	G Index
Down Forecast	Lagged Momentum	Leverage Component of B/P	Gross Profitability
Exchange Switch	Long-term Reversal	Marketing/MV	G-Score
Growth in Inventory	Max	Org. Capital	G-Score 2
Growth in LTNOA	Momentum	R&D/MV	Herfindahl
IPO	Momentum and Long-term Reversal	Sales/Price	Investment
IPO + Age	Momentum-Ratings		Leverage
IPO no R&D	Momentum-Reversal		M/B and Accruals
Mergers	Price		NOA
Post Earnings Drift	Seasonality		Operating Leverage
R&D Increases	Short Interest		O-Score
Ratings Downgrades	Short-term Reversal		Pension Funding
Repurchases	Size		Percent Operating Accrual
Revenue Surprises	Volume		Percent Total Accrual
SEOs	Volume Trend		Profit Margin
Share Issuance 1-Year	Volume Variance		Profitability
Share Issuance 5-Year	Volume-Momentum		ROE
Spinoffs	Volume/MV		Sales Growth
Sustainable Growth			Tax
Total External Finance			Z-Score
Up Forecast			
ΔCapex - ΔIndustry CAPEX			
ΔNoncurrent Op. Assets			
ΔSales - ΔInventory			
ΔSales - ΔSG&A			
ΔWork. Capital			

Returns post-sample and post-publication

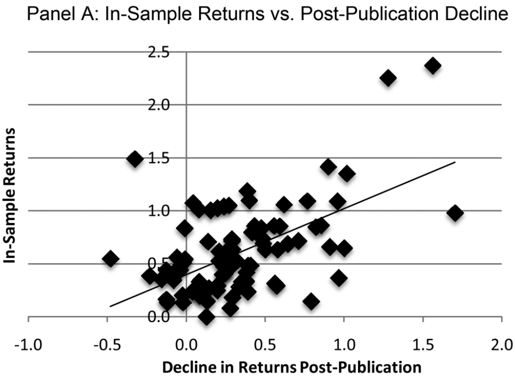
- The main panel regression of the paper is specified as follows (where α_i captures predictor fixed effects)

$$R_{it} = \alpha_i + \beta_1 \text{Post sample dummy}_{it} + \beta_2 \text{Post publication dummy}_{it} + e_{it} \quad (1)$$

Variables	(1)	(2)	(3)	(4)
Post-Sample (S)	−0.150***	−0.180**	0.157	0.067
	(0.077)	(0.085)	(0.103)	(0.112)
Post-Publication (P)	−0.337***	−0.387***	−0.002	−0.120
	(0.090)	(0.097)	(0.078)	(0.114)
S × Mean			−0.532***	
			(0.221)	
P × Mean			−0.548***	
			(0.178)	
S × t-statistic				−0.061***
				(0.023)
P × t-statistic				−0.063***
				(0.018)
Predictor FE?	Yes	Yes	Yes	Yes
Observations	51,851	45,465	51,851	51,944
Predictors (N)	97	85	97	97
Null : S = P	0.024	0.021		
Null: P = −1 × (mean)	0.000	0.000		
Null: S = −1 × (mean)	0.000	0.000		

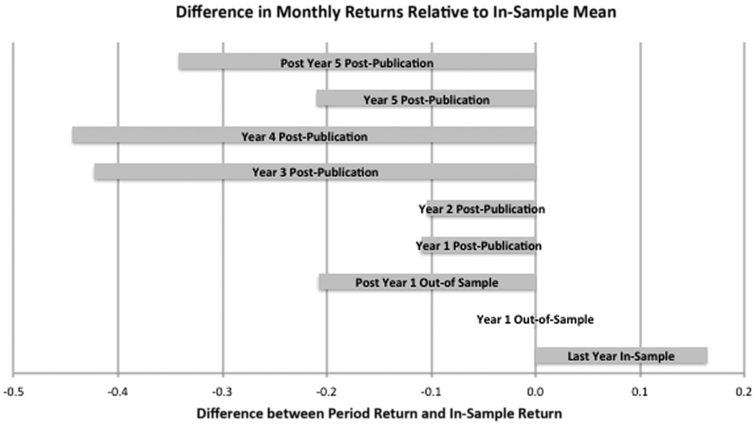
- We have reductions in returns of $15/58.2 = 26\%$ and $33.7/58.2 = 58\%$, respectively, relative to the average in-sample return of 58.2bps

- Higher in-sample returns are correlated with larger reductions in post-publication returns (similar for t-statistics)



Returns around sample-end and publication dates

- We then examine changes in predictability by considering a set of partitions of the sample in event time at higher granularity



Time trends and persistence in returns

- Are the results due to time trends, persistencies, or lower costs of *corrective* trading instead of academic research?

- The results are not supportive of this hypothesis

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Time	−0.069***		−0.069***			
	(0.011)		(0.026)			
Post-1993		−0.120	0.303***			
		(0.074)	(0.118)			
Post-Sample			−0.190**	−0.179**	−0.132*	−0.128
			(0.081)	(0.080)	(0.076)	(0.078)
Post-Publication			−0.362***	−0.310**	−0.295***	−0.258***
			(0.124)	(0.122)	(0.089)	(0.093)
1-Month Return					0.114***	
					(0.015)	
12-Month Return						0.020***
						(0.004)
Observations	51,851	51,851	51,851	51,851	51,754	50,687
Char. FE?	Yes	Yes	Yes	Yes	Yes	Yes
Time FE?	No	No	No	Yes	No	No

Returns across predictor types

- To test for differences among predictor groups, we can run the following regression model

$$R_{it} = \alpha_i + \beta_1 \text{Post publication dummy}_{it} + \beta_2 \text{Predictor type dummy}_i + \beta_3 \text{Post publication dummy}_{it} \times \text{Predictor type dummy}_i + e_{it} \quad (2)$$

Variable	(1)	(2)	(3)	(4)
Post-Publication (P)	−0.208***	−0.316***	−0.310***	−0.301***
	(0.059)	(0.097)	(0.080)	(0.089)
Market	0.304***			
	(0.079)			
P × Market	−0.244			
	(0.169)			
Event		−0.098**		
		(0.046)		
P × Event		0.105		
		(0.091)		
Valuation			−0.056	
			(0.063)	
P × Valuation			0.186	
			(0.131)	
Fundamental				−0.201***
				(0.045)
P × Fundamental				0.025
				(0.089)
Constant	0.482***	0.606***	0.585***	0.630***
	(0.036)	(0.052)	(0.000)	(0.053)
Observations	51,851	51,851	51,851	51,851
Predictors	97	97	97	97
Type + (P × Type)	0.060	0.007	0.121	−0.176
p-value	0.210	0.922	0.256	0.012

- The results so far are consistent with the idea that publication attracts arbitrageurs, which leads to lower returns post-publication
- Costly arbitrage can prevent mispricing from being fully eroded, which suggests that predictor portfolios concentrated in stocks that are costlier to arbitage (e.g., smaller stocks, illiquid stocks, and high idiosyncratic risk stocks) should decline less post-publication
- McLean and Pontiff (2016) consider three transaction costs variables (size, bid-ask spreads, dollar volume) and two holding cost variables (idiosyncratic risk, dividend-payer dummy) and run the following regression model

$$\begin{aligned} R_{it} = \alpha_i + \beta_1 \text{Post publication dummy}_{it} + \beta_2 \text{Arbitrage cost}_i \\ + \beta_3 \text{Post publication dummy}_{it} \times \text{Arbitrage cost}_i + e_{it} \end{aligned} \tag{3}$$

Costly arbitrage results

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Post-Pub. (P)	−0.190	−0.139	0.215	−0.242	−0.321	−0.264**
	(0.274)	(0.235)	(0.230)	(0.273)	(0.211)	(0.078)
P × Size	−0.138					
	(0.459)					
Size	−1.064**					
	(0.236)					
P × Spreads		−0.301				
		(0.603)				
Spreads		1.228**				
		(0.252)				
P × Dol.Vol.			−1.059*			
			(0.500)			
Dol. Vol.			0.215			
			(0.308)			
P × Idio. Risk				−0.047		
				(0.554)		
Idio. Risk				2.064***		
				(0.330)		
P × Div.					−0.321	
					(0.211)	
Div.					−0.526***	
					(0.145)	
P × Index						−0.009
						(0.019)
Index						−0.056***
						(0.011)
Constant	1.145***	0.146*	0.476***	−0.469***	0.855***	0.565***
	(0.130)	(0.174)	(0.144)	(0.171)	(0.097)	(0.000)
Observations	51,851	51,851	51,851	51,851	51,851	51,851
CA + (P × CA)	−1.202	0.927	−0.844	2.017	−0.847	−0.065
p-value	0.003	0.096	0.000	0.000	0.144	0.000

Post-sample and -publication trading activity dynamics

- If academic publication provides market participants with information, then informed trading activity should affect indicators of trading beyond prices
- Below, we investigate whether variance, trading volume, dollar trading volume, and short interest increase in predictor portfolios post-publication
- Note that a significant post-sample coefficient is indicative of the informational content being out prior to publication (e.g., working papers and conferences)

Variables	Variance	Trading volume	Dollar volume	Short-long short interest
Post-Sample (S)	−0.054***	0.092***	0.066***	0.166***
	(0.007)	(0.001)	(0.007)	(0.014)
Post-Publication (P)	−0.065***	0.187***	0.097***	0.315***
	(0.008	(0.013)	(0.007)	(0.013)
Observations	52,632	52,632	52,632	41,026
Time FE?	Yes	Yes	Yes	No
Predictor FE?	Yes	Yes	Yes	Yes
Null: S = P	0.156	0.000	0.000	0.000

Correlation structure and publication

- If predictor returns reflect mispring, and if mispricing has a common source, then we should expect that
  1. In-sample returns are significantly related to other in-sample portfolios
  2. Once a predictor is published (and attracts arbitrageurs), its return should correlate (less) more with the returns of other (pre-) post-publication predictors
- We examine this by regressing predictor portfolio returns on the returns of an
  1. equal-weighted index of all other portfolios that are pre-publication
  2. equal-weighted index of all of the other portfolios that are post-publication
  3. A post-publication dummy and interactions with the indices

Variables	Coefficients
In-Sample Index Returns	0.748***
	(0.027)
Post-Publication Index Return	−0.008
	(0.004)
P × In-Sample Index Returns	−0.674***
	(0.033)
P × Post-Publication Index Return	0.652***
	(0.045)
Publication (P)	−0.081
	(0.042)
Constant	0.144***
	(0.019)
Observations	42,975
Predictors	97

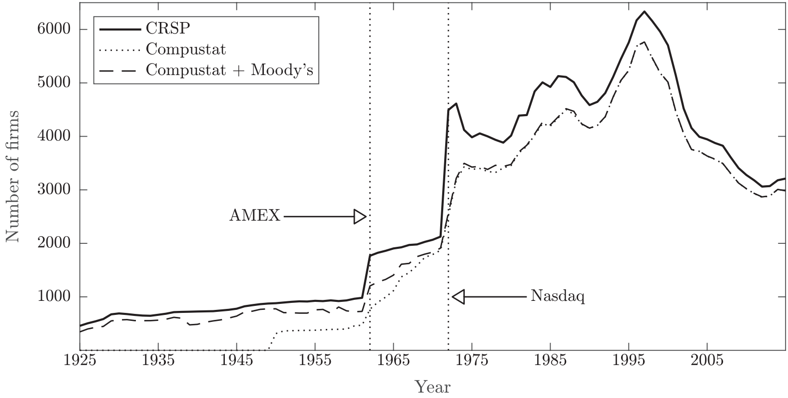
Linnainmaa and Roberts (2018)

Linnainmaa and Roberts (2018)

- Linnainmaa and Roberts (2018) ask the following question: Does accounting-based anomalies persist out-of-sample (both pre- and post-sample)?
- To examine this, they consider a broad selection of 36 accounting-based anomalies over three distinct periods
  1. "In-sample" denotes the sample frame used in the original discovery of an anomaly
  2. "Pre-sample" denotes the sample frame occurring prior to the in-sample period
  3. "Post-sample" denotes the sample frame occurring after the in-sample period
- Similar to McLean and Pontiff (2016), they distinguish between three possible explanations
  1. Unmodeled risk $\Rightarrow$ Stocks risk are multidimensional and models are misspecified
  2. Mispricing $\Rightarrow$ Investor irrationality and limits to arbitrage cause anomaly returns
  3. Data-snooping $\Rightarrow$ Anomalies are artifacts of chance error
- **Main conclusion:** Most anomaly returns are decidedly an in-sample phenomenon consistent with data-snooping (i.e., most risk factors are false)

Number of firms available

- One of the main contribution of Linnainmaa and Roberts (2018) is to back-fill accounting information using Moody's Industrial and Railroad manuals to obtain out-of-sample data points back in time
- See also Baltussen et al. (2021) for a paper constructing a pre-CRSP dataset



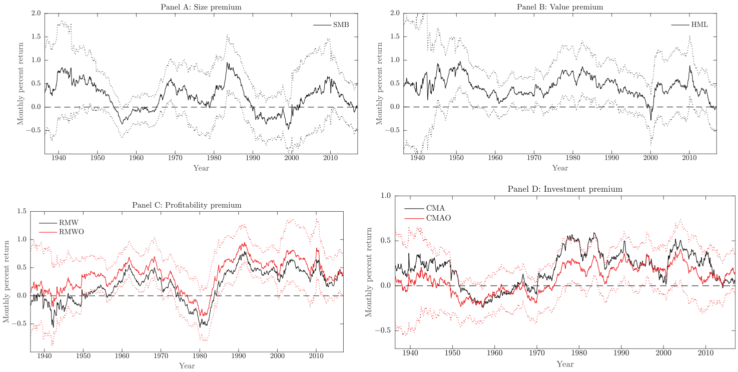
Case study: Fama-French five factors

- To illustrate their approach, consider the case of the Fama and French (2015) five-factor model using pre-1963 accounting data

Panel A: Monthly percent returns

Portfolio	Pre-1963 sample 1926:7 –1938:6	Pre-1963 sample 1938:7 –1963:6	Pre-1963 sample 1926:7 –1963:6	1963:7 –2016:12
	Portfolios sorted by size and book-to-market			
Small Growth	0.73	1.17	1.03	0.88
Neutral	1.04	1.32	1.23	1.29
Value	1.41	1.62	1.55	1.40
Big Growth	0.80	0.99	0.93	0.88
Neutral	0.73	1.16	1.02	0.96
Value	0.88	1.51	1.31	1.13
Size factor	0.26	0.15	0.18	0.20
	(0.62)	(1.05)	(1.11)	(1.54)
Value factor	0.38	0.48	0.45	0.38
	(0.66)	(2.87)	(2.06)	(3.40)
	Portfolios sorted by size and profitability			
Small Weak	0.88	1.32	1.18	0.96
Neutral	0.78	1.38	1.19	1.24
Robust	0.75	1.37	1.17	1.31
Big Weak	0.81	1.09	1.00	0.83
Neutral	0.76	1.10	0.99	0.90
Robust	0.68	1.23	1.05	1.03
Profitability factor	−0.13	0.09	0.02	0.28
	(−0.31)	(0.64)	(0.14)	(3.09)
	Portfolios sorted by size and investment			
Small Aggressive	0.84	1.23	1.10	0.98
Neutral	1.58	1.36	1.43	1.34
Conservative	1.06	1.36	1.27	1.33
Big Aggressive	0.69	1.11	0.97	0.90
Neutral	1.14	1.12	1.13	0.94
Conservative	0.86	1.04	0.98	1.08
Investment factor	0.19	0.03	0.09	0.26
	(0.79)	(0.32)	(0.80)	(3.28)

- Consider the ten-year rolling window average returns to the Fama and French (2015) factors
- See also Fama and French (2021) for a discussion of the recent performance of the value premium



Risk factors under consideration

- The 36 accounting-based anomalies under consideration, and their groupings, are specified below and information about the construction can be found in the paper appendix

Category	No.	Anomaly	Original study	Original sample
Profitability	1	Gross profitability	Novy-Marx (2013)	1963–2010
	2	**Operating profitability***	Fama and French (2015)	1963–2013
	3	Return on assets*	Haugen and Baker (1996)	1979–1993
	4	Return on equity*	Haugen and Baker (1996)	1979–1993
	5	Profit margin	Soliman (2008)	1984–2002
	6	Change in asset turnover	Soliman (2008)	1984–2002
Earnings quality	7	Accruals*	Sloan (1996)	1962–1991
	8	Net operating assets	Hirshleifer, Hou, Teoh, and Zhang (2004)	1964–2002
	9	Net working capital changes	Soliman (2008)	1984–2002
Valuation	10	**Book-to-market**	Fama and French (1992)	1963–1990
	11	Cash flow / price	Lakonishok, Shleifer, and Vishny (1994)	1968–1990
	12	Earnings / price	Basu (1977)	1957–1971
	13	Enterprise multiple*	Loughran and Wellman (2011)	1963–2009
	14	Sales / price	Barbee, Mukherji, and Raines (1996)	1979–1991
Investment and growth	15	**Asset growth**	Cooper, Gulen, and Schill (2008)	1968–2003
	16	Growth in inventory	Thomas and Zhang (2002)	1970–1997
	17	Sales growth	Lakonishok, Shleifer, and Vishny (1994)	1968–1990
	18	Sustainable growth	Lockwood and Prombutr (2010)	1964–2007
	19	Adjusted CAPX growth*	Abarbanell and Bushee (1998)	1974–1993
	20	Growth in sales − inventory	Abarbanell and Bushee (1998)	1974–1993
	21	Investment growth rate*	Xing (2008)	1964–2003
	22	Abnormal capital investment*	Titman, Wei, and Xie (2004)	1973–1996
	23	Investment to capital*	Xing (2008)	1964–2003
	24	Investment-to-assets	Lyandres, Sun, and Zhang (2008)	1970–2005
Financing	25	Debt issuance*	Spiess and Affleck-Graves (1999)	1975–1994
	26	Leverage	Bhandari (1988)	1948–1979
	27	One-year share issuance	Pontiff and Woodgate (2008)	1970–2003
	28	Five-year share issuance	Daniel and Titman (2006)	1968–2003
	29	Total external financing*	Bradshaw, Richardson, and Sloan (2006)	1971–2000
Distress	30	O-score	Dichev (1998)	1981–1995
	31	z-score*	Dichev (1998)	1981–1995
	32	Distress risk	Campbell, Hilscher, and Szilagyi (2008)	1963–2003
Other	33	Industry concentration	Hou and Robinson (2006)	1951–2001
Composite anomalies	34	Piotroski's F-score	Piotroski (2000)	1976–1996
	35	M/B and accruals*	Bartov and Kim (2004)	1981–2000
	36	QMJ: Profitability	Asness, Frazzini, and Pedersen (2013)	1956–2012

Returns in pre-, in-, and post-sample periods

- The main result of Linnainmaa and Roberts (2018) is that anomaly returns are a decidedly in-sample phenomenon, which is clearly evident in their Table 6

Measure	Pre-sample	In-sample	Post-sample	Differences: Pre − In	Differences: Post − In	Differences: Post − Pre
Panel A: Full pre-1963 sample						
Average returns						
Average	0.08	0.29	0.09	−0.21	−0.20	0.00
return	(2.21)	(7.01)	(1.72)	(−3.78)	(−3.69)	(0.03)
Sharpe	0.15	0.54	0.13	−0.39	−0.42	−0.03
ratio	(3.38)	(7.57)	(1.52)	(−4.71)	(−4.14)	(−0.30)
CAPM						
Alpha	0.15	0.34	0.17	−0.20	−0.18	0.02
	(4.80)	(9.75)	(3.50)	(−4.27)	(−3.44)	(0.38)
Information	0.22	0.66	0.27	−0.43	−0.40	0.04
ratio	(5.08)	(9.72)	(2.99)	(−5.43)	(−3.83)	(0.43)
Three-factor model						
Alpha	0.17	0.27	0.12	−0.10	−0.15	−0.05
	(6.42)	(10.12)	(3.19)	(−2.57)	(−3.44)	(−1.10)
Information	0.28	0.60	0.25	−0.32	−0.35	−0.03
ratio	(6.35)	(9.91)	(2.86)	(−4.26)	(−3.46)	(−0.32)

Effect of state date on returns

- To investigate the sensitivity to the in-sample period, Linnainmaa and Roberts (2018) run a panel regression of the form

$$\text{anomaly}_{it} = \beta_0 + \beta_1 \text{Pre-sample}_{it} + \mu_i + \varepsilon_{it} \tag{4}$$

where μ_i is an anomaly fixed effect and Pre-sample_{it} is an indicator equal to one if the anomaly-month observation falls in the time period before the start date of the anomaly's in-sample start date

	Average return		CAPM alpha		FF3 alpha		
Start year	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	No. of obs.
1963	0.30	−0.15	0.36	−0.18	0.27	−0.14	14,793
	(6.77)	(−2.16)	(10.07)	(−2.97)	(10.35)	(−3.18)	
1964	0.30	−0.15	0.36	−0.19	0.28	−0.13	14,385
	(6.77)	(−1.93)	(10.06)	(−2.86)	(10.40)	(−2.68)	
1965	0.30	−0.13	0.36	−0.17	0.28	−0.11	13,977
	(6.78)	(−1.58)	(10.07)	(−2.42)	(10.40)	(−2.27)	
1966	0.30	−0.13	0.37	−0.17	0.28	−0.12	13,569
	(6.84)	(−1.46)	(10.15)	(−2.26)	(10.45)	(−2.21)	
1967	0.31	−0.09	0.37	−0.13	0.29	−0.12	13,161
	(6.89)	(−0.98)	(10.21)	(−1.84)	(10.50)	(−2.25)	
1968	0.31	−0.07	0.37	−0.13	0.29	−0.13	12,753
	(6.90)	(−0.80)	(10.20)	(−2.07)	(10.45)	(−2.31)	
1969	0.31	−0.11	0.38	−0.17	0.29	−0.18	12,345
	(6.77)	(−1.21)	(10.09)	(−2.53)	(10.44)	(−3.04)	
1970	0.30	−0.22	0.37	−0.24	0.29	−0.25	11,937
	(6.53)	(−2.68)	(9.91)	(−3.37)	(10.22)	(−4.12)	
1971	0.31	−0.22	0.37	−0.26	0.29	−0.28	11,532
	(6.63)	(−2.65)	(9.88)	(−3.78)	(10.02)	(−4.60)	
1972	0.32	−0.21	0.38	−0.26	0.29	−0.31	11,136
	(6.64)	(−2.18)	(9.86)	(−3.24)	(9.86)	(−4.52)	
1973	0.31	−0.24	0.38	−0.28	0.31	−0.27	10,740
	(6.38)	(−2.20)	(9.70)	(−3.31)	(10.17)	(−3.51)	

Volatility across the eras

- Linnainmaa and Roberts (2018) argue that data-snooping works through t-values, so examining volatility across the three periods may be worthwhile
- Consider the panel regression specification for squared demeaned returns

$$(\text{anomaly}_{it} - \bar{r}_{it})^2 = \beta_0 + \beta_1 \text{In-sample}_{it} + \beta_2 \text{Post-sample}_{it} + \mu_{it} + \varepsilon_{it} \tag{5}$$

- Estimating the model yields the following parameter estimates
  - $\beta_1 = -0.61$ with a t-stat of -2.88
  - $\beta_2 = 0.38$ with a t-stat of 0.79

  indicating that volatility is lower during the in-sample period, which is consistent with data-snooping contaminating the distribution of in-sample returns

Changes in correlation structure of returns

- Finally, we can investigate effects upon the correlation structure as in McLean and Pontiff (2016) by running the regression

$$\begin{aligned}\text{anomaly}_{it} &= a + b_1\text{in-sample index}_{-i,t} + b_2\text{post-sample index}_{-i,t} + b_3\text{post}_{it} \\ &\quad + \text{post}_{it} \times \left(b_4\text{in-sample index}_{-i,t} + b_5\text{post-sample index}_{-i,t}\right) + e_{i,t}\end{aligned}$$

Regressor	Coefficient	t-value
Regression 1: In-sample versus post-sample anomalies		
Intercept	0.05	4.54
Main effects		
In-sample index$_{-i,t}$	0.74	33.98
Post-sample index$_{-i,t}$	0.08	7.46
Post$_{i,t}$	−0.06	−2.23
Interactions		
Post$_{i,t}$ × In-sample index$_{-i,t}$	−0.53	−13.74
Post$_{i,t}$ × Post-sample index$_{-i,t}$	0.46	11.19
Adjusted R^2		17.9%
N		15,152
Regression 2: In-sample versus pre-sample anomalies		
Intercept	0.07	4.35
Main effects		
In-sample index$_{-i,t}$	0.74	28.90
Pre-sample index$_{-i,t}$	0.07	3.42
Pre$_{i,t}$	−0.04	−2.09
Interactions		
Pre$_{i,t}$ × In-sample index$_{-i,t}$	−0.69	−22.72
Pre$_{i,t}$ × Pre-sample index$_{-i,t}$	0.48	13.68
Adjusted R^2		9.3%
N		13,650

Other perspectives in the debate

Factors are mostly false

- There is a large camp, mainly headed by Campbell R. Harvey, that claims that most factors are false and that there is a pronounced replication crisis in finance (Harvey et al., 2016, Harvey, 2017, Harvey and Liu, 2019, 2021)
  - Hou et al. (2020) re-evaluate 452 anomalies and find that mitigating for microcaps using NYSE breakpoints and value-weighted returns leads to a failure rate of 65% using a 1.96 cutoff. It increases to a failure rate of 82% using a cutoff of 2.78
  - Tian (2020) runs a data-mining experiment in which she randomly constructs hundreds of three-factor models. She finds that many outperform well known models from the literature, including those with four and five factor. This suggest that the threshold of factor model success needs to be raised
  - Chordia et al. (2020) use information from over 2 million randomly generated trading strategies (using real data and strategies that survive the publication process) to infer the statistical properties of factor strategies. They compute t-statistic threshold that control for multiple hypothesis testing at 3.8 and 3.4 for time-series and cross-sectional regressions, respectively, Failing to account for multiple hypothesis testing leads to a false rejection about 45% of the time

Factors are mostly true

- There is similarly a growing literature arguing that most factors are true and can be successfully replicated
  - Engelberg et al. (2018) find that anomaly returns are 50% higher on corporate news days and six times higher on earnings announcement days. They argue that the results point to the idea that anomaly returns are driven by biased expectations that are (partly) corrected by information arrivals
  - Jacobs and Müller (2020) extend the work of McLean and Pontiff (2016) by investigating 241 anomalies in 39 countries. Their results similar point to mispricing rather than data mining
  - Calluzzo et al. (2019) show that there is an increase in anomaly-related trading when information about the anomalies is readily available through academic publications
  - Chen and Zimmermann (2020) argue that bias-adjusted returns are only 12.3% smaller than in-sample returns, which is well within the McLean and Pontiff (2016) upper bounds and points to mispricing
  - Chen (2021a,b) provides thought experiments and alternative views on the results of Harvey et al. (2016) and argues that p-hacking cannot be as widely applied as argued and that most factors are true as a consequence
  - Jensen et al. (2021) challenge the dire view of finance research. They develop and estimate a Bayesian model of factor replication, which leads to different conclusions. They find a baseline replication rate as high as 55.6%! (plus make available a new global dataset at https://www.bryankellyacademic.org under Data)

Time-varying factor returns

- The entire debate so far, has taken an unconditional view; meaning that we do not take into account that many of the factors are related to the business cycle
- Kelly and Pruitt (2013) find that bm ratios predicts momentum, size, and industry portfolios
- Baba Yara et al. (2021) find that the value spread predicts the return value strategies across many different assets
- Haddad et al. (2020) show that common components of anomalies are predictable (OoS) using their value spread $\rightarrow$ exploitable for factor timing!
- Smith and Timmermann (2022) find "breaks" in the risk premium of many anomalies during times of economic turbulence

Factor momentum

- More recently, a couple of papers has examined factor momentum
- Avramov et al. (2017) show that sorting factors on most recent performance is highly profitable
- Ehsani and Linnainmaa (2022) show that the momentum factor can be fully explained by factor momentum (TS momentum)
- Arnott et al. (2023) show that cross-sectional factor momentum has high alphas

➜ Factors seems to have autocorrelation that you can exploit!

Open source asset pricing

- Last, I want to direct attention to a great initiative and a compelling voice in the debate about replicability in finance: https://www.openassetpricing.com
- Chen and Zimmermann (2021) provide open source code, characteristics data, and portfolio returns for about 200 (204 to be precise) factor portfolios
- This is a great source of data for re-examining cross-sectional predictors by combining their data with the CRSP database (you can do that using PERMNO)

Let us examine short term factor momentum!

- We consider the (scaled) characteristics from Chen and Zimmermann (2021) which we merge with our CRSP dataset
- Data from 1986 to 2019
- Follow the recommendations of Hou et al. (2020), i.e. NYSE-based breakpoints, VW portfolios, and consider quintile portfolios.
- Each factor is defined as long in portfolio 5 and short portfolio 1
- Some of the factors in Chen and Zimmermann (2021) are based on discrete chacteristics (such as the sin factor). We will ignore these
- We then sort our CS of factors based on most recent performance (t-1 return) into 5 EW portfolios

Short term factor momentum is highly profitable!

- The five factor portfolios generates the following average returns:

	1	2	3	4	5	5-1
Average returns	-0.36	1.53	3.07	4.09	8.53	8.89
SR	-0.04	0.33	1.07	0.89	0.85	0.47
t-stat	-0.23	1.73	5.78	5.50	6.05	3.25

➜ Factor performance is an increasing function of most recent factor performance!

Aaand over time

➜ Factor momentum has disappreared since 2010... Why?

Potential projects

Potential projects

- Investigate the returns and pricing ability pre- and post-publication to anomalies/factors
- Investigate whether there are publication effects in other asset classes (e.g., currencies or bonds)
- Combine a CRSP-based portfolio sorts or a cross-sectional asset pricing exercise with robustness checks inspired by this literature (e.g., re-examine with stricter thresholds)
- Test whether selected time-series predictors similarly display publication effects inspired by this literature
- Re-examine selected risk factors/anomalies out-of-sample and investigated whether results can be replicated and extended

References

Arnott, R. D., M. Clements, V. Kalesnik, and J. T. Linnainmaa (2023). Factor momentum. *Review of Financial Studies*.

Avramov, D., S. Cheng, A. Schreiber, and K. Shemer (2017). Scaling up market anomalies. *The Journal of Investing 26*(3), 89–105.

Baba Yara, F., M. Boons, and A. Tamoni (2021). Value return predictability across asset classes and commonalities in risk premia. *Review of Finance 25*(2), 449–484.

Baltussen, G., B. P. van Vliet, and P. van Vliet (2021). The cross-section of stock returns before 1926 (and beyond). Working paper, Erasmus School of Economics.

Calluzzo, P., F. Moneta, and S. Topaloglu (2019). When anomalies are publicized broadly, do institutions trade accordingly? *Management Science 65*(10), 4451–4949.

Chen, A. Y. (2021a). The limits of p-hacking: Some thought experiments. *Journal of Finance 76*(5), 2447–2480.

Chen, A. Y. (2021b). Most claimed statistical findings in cross-sectional return predictability are likely true. Working paper.

Chen, A. Y. and T. Zimmermann (2020). Publiation bias and the cross-section of stock returns. *Review of Asset Pricing Studies 10*(2), 249–289.

Chen, A. Y. and T. Zimmermann (2021). Open source cross sectional asset pricing. *Critical Finance Review*, Forthcoming.

Chordia, T., A. Goyal, and A. Saretto (2020). Anomalies and false rejections. *Review of Financial Studies 33*(5), 2134–2179.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance 66*(5), 1047–1108.

Ehsani, S. and J. T. Linnainmaa (2022). Factor momentum and the momentum factor. *The Journal of Finance 77*(3), 1877–1919.

Engelberg, J., R. D. McLean, and J. Pontiff (2018). Anomalies and news. *Journal of Finance 73*(5), 1971–2001.

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 K. R. French (2021). The value premium. *Review of Asset Pricing Studies 11*(1), 105–121.

Haddad, V., S. Kozak, and S. Santosh (2020). Factor timing. *The Review of Financial Studies 33*(5), 1980–2018.

Harvey, C. R. (2017). Presidential address: The scientific outlook in financial economics. *Journal of Finance 72*(4), 1399–1440.

Harvey, C. R. and Y. Liu (2019). A census of the factor zoo. Working paper, Duke University.

Harvey, C. R. and Y. Liu (2021). Uncovering the iceberg from its tip: A model of publication bias and p-hacking. Working paper, Duke University.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

Hou, K., C. Xue, and L. Zhang (2020). Replicating anomalies. *Review of Financial Studies 33*(5), 2019–2133.

Jacobs, H. and S. Müller (2020). Anomalies across the globe: One public, no longer existent? *Journal of Financial Economics 153*, 213–230.

Jensen, T. I., B. Kelly, and L. H. Pedersen (2021). Is there a replication crisis in finance? *Journal of Finance*, Forthcoming.

Kelly, B. and S. Pruitt (2013). Market expectations in the cross-section of present values. *The Journal of Finance 68*(5), 1721–1756.

Linnainmaa, J. T. and M. R. Roberts (2018). The history of the cross-section of stock returns. *Review of Financial Studies 31*(7), 2606–2649.

McLean, R. D. and J. Pontiff (2016). Does academic research destroy stock return predictability? *Journal of Finance 71*(1), 5–32.

Smith, S. C. and A. Timmermann (2022). Have risk premia vanished? *Journal of Financial Economics 145*(2), 553–576.

Tian, M. (2020). Firm characteristics and empirical factor models: A model mining experiment. *Review of Financial Studies*, Forthcoming.

ML Predictability

Empirical Asset Pricing

Mads Markvart Kjær

Department of Economics and Business Economics, Aarhus University, CREATES

E-mail: mads.markvart@econ.au.dk

Spring 2022

What is Machine Learning (ML)?

Definition 1 (Samuel): Machine learning

Machine learning is the use of algorithms and statistical models that computer systems employ to perform a specific task without using explicit instructions. It relies on patterns and inference instead. Machine learning algorithms build a mathematical model based on sample data, known as “training data", in order to make predictions or decisions without being explicitly programmed to perform the task (Samuel, 1959).

Definition 2 (Mitchell): Machine learning

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E (Mitchell, 1997).

Another definition (sort of)

> Interviewer: What's your biggest strength?
>
> Me: I'm a fast learner.
>
> Interviewer: What's 11 * 11?
>
> Me: 65.
>
> Interviewer: Not even close. It's 121.
>
> Me: It's 121.

Another definition (sort of)

A different way of thinking

- Previos in the course, we started from a model (i.e., a DGP) and our outcome was to estimate the true parameters (like the case of the SDF)
- Now, we move into the area in which we (in the extreme case) do not care about the DGP but ultimately want to do predictions!

ML in Empiricial Asset Pricing

ML in empirical asset pricing

Definition: Machine learning in practice

Machine learning is:

- A diverse collection of high-dimensional models for statistical prediction
- Combined with so-called regularization methods for model selection and the mitigation of overfit
- Efficient algorithms for searching among a vast number of potential model specifications.

ML in empirical asset pricing

- Element i) enhances the flexibility relative to more traditional econometric techniques - brings hope of better approximating the unknown and likely complex (true) data generating process.
- Element ii) guards against overfitting that is likely to arise due to the higher flexibility (and no. of parameters) and low signal-to-noise environment.
- Element iii) describes ML techniques that are designed to approximate an optimal specification with manageable computational costs.

$\rightarrow$ For the following three reasons, ML is well-suitable for empirical asset pricing!

Reason 1: The fundamental research agenda

- Empirical asset pricing centres around estimating the (conditional) expectation of asset returns (recall the course introduction):
  1. Time series dynamics of assets' risk premium (excess returns).
  2. Cross-sectional variation in assets' average excess returns.
- Both are fundamentally a question of predictability either in the time series or the cross-sectional direction.
- Which requires a specification/search or learning the true function for the conditionally expected return.

$\Rightarrow$ ML takes the view that the best approximation model for expected returns should be learned!

Reason 2: The collection of relevant variables is enormous

- The profession has accumulated a staggering list of predictors and risk factors that has been argued are important for prediction/measuring assets' excess returns.

Snip of relevant variables

- Harvey et al. (2016) have identified 316 risk factors.
- Green et al. (2017) have collected 94 stock-level predictive signals.
- Welch and Goyal (2008) have collected 20 (at various frequencies) predictors for the aggregate market risk premium.
- McCracken and Ng (2016) have collected in excess of 100 macroeconomic variables that are often used for predicting stock returns, both in the time series and in the cross-section.

...

Reason 2: The collection of relevant variables is enormous

- Those variables are often correlated and do likely not represent in excess of, say, 300 orthogonal and important signals.
- Traditional econometric techniques would fail massively (and not even be feasible) in those cases.

$\Rightarrow$ ML techniques are well suited for so-called high-dimensional settings, by encompassing variable selection and dimension reduction techniques.

Reason 3: The functional form is unknown and likely complex

- Further complicating the problem of empirical asset pricing is a fundamental ambiguity regarding the functional form through which the high-dimensional predictors should enter into excess returns.
- Should they enter linearly? Non-linearly? If so, how? Through interactions? In squared or cubed forms?
- This radically proliferates the number of predictors and the dimensionality of the problem!
- There is not much help in the literature of theoretical asset pricing in which it all comes down to an assumption...

Reason 3: The functional form is unknown and likely complex

$\Rightarrow$ ML techniques are well suited for this, as it learns, with an emphasis on parsimony, the best specifications due to

1. considering several specifications,
2. incorporating non-linear models
3. penalizing additional parameters and using conservative model selection to avoid overfitting

What ML cannot do

- Fundamentally, ML has great potential for improving the estimation of the (conditional) expectation of excess returns, that is, best approximate the form for $\mathbb{E}[r^e_{it+1}|\mathcal{F}_t]$.
- But this is just *measurements*. It is silent about economic mechanisms and equilibria.
- This requires putting structure on the problem and applying ML in an appropriate manner.
- The approach in Gu et al. (2020) is not generally about this, but we will address this a bit here anyway, and otherwise much more in the last topic of the course on cross-sectional asset pricing.

A nice overview!

- Check out Masini et al. (2020) for an overview of ML methods in time series forecasting.

ML Methods

Return predictability via ML

Return predictability in very general terms

In its most general form, assets' excess returns can always be expressed as

$$r^e_{i,t+1} = \mathbb{E}_t[r^e_{i,t+1}] + \varepsilon_{i,t+1}, \tag{1}$$

where

$$\mathbb{E}_t[r^e_{i,t+1}] = g^*(z_{i,t}), \tag{2}$$

$\varepsilon_{i,t+1}$ is a forecast error term and z_{it} is a P-dimensional set of (possibly stock-specific) predictors, $i = 1, \ldots, N$, and $t = 1, \ldots, T$.

- The objective is to learn the best specification of the general function $g(\cdot)$, and we will later see numerous ML techniques that are quite successful in this.
- Note that this function is independent of i and t.

Return predictability via ML (in practice)

- The main focus in this lecture will be on out-of-sample predictability.

Out-of-sample predictability

* Assessing predictability in-sample entails estimating the predictive regression using the full range of available observations
* Assessing out-of-sample predictability, conversely, entails using information available at time t only to forecast returns at time $t+1$
* To emulate a forecaster in real-time, we split the total sample (T) in two parts: in-sample (initial) and out-of-sample

$$\underbrace{t = 1, 2, \ldots, R}_{\text{In-sample}}, \underbrace{R+1, R+2, \ldots, T}_{\text{Out-of-sample}} \tag{41}$$

* One can either estimate the regression coefficients using a rolling or an expanding window of data (benefits/drawback?)
* Irrespective of choice, we end up with a sequence of forecasts $\{\widehat{r}_i\}_{i=R+1}^{T}$ and forecast errors $\{\widehat{\varepsilon}_i\}_{i=R+1}^{T}$ for evaluation

Hyperparameters

- All methods we will go through depends on one (or more) choice (s) made by the researcher
- Many of these choices goes under the definition of hyperparameters:

Definition: Hyperparameters

Most of the ML techniques depends on parameters that are not part of the model, but rather specifies parts of the estimation procedure.

- These typically control e.g. regularization (a defense against overfitting).
- Examples include penalty parameters in LASSO or number of random trees in a forest.

Choosing hyperparameters

For instance, the following three approaches are common

- Cross-validation
- Information criteria (Related to model selection and regularization)
- They can be “tuned" to optimize out-of-sample performance

  $\rightarrow$ We will for now mostly focus on the tuning approach

Return predictability via ML (in practice)

- We need to re-define the design of the out-of-sample study by introducing additional sample splitting and hyperparameters (or tuning parameters).

Definition: Sample splitting for tuning hyperparameters

We can tune hyperparameters by splitting the sample into three disjoint (not the usual two) time periods that maintain the temporal ordering of data.

1. Training period: This period/data is used to estimate (train) the model, given a choice of hyperparameters.
2. Validation period: The trained model from the training period is then used to form predictions for the validation period. Here, the objective function from the forecast evaluation is computed (e.g. out-of-sample R^2).
3. Testing period: This sample period is not used in estimation/training, nor validation step, and is truly out-of-sample, used to evaluate the methods' predictive ability.

$$\underbrace{1,2,\ldots,R_1}_{\text{training}},\underbrace{R_1+1,R_1+2,\ldots,R_2}_{\text{validation}},\underbrace{R_2+1,R_2+2,\ldots,T}_{\text{testing}} \tag{3}$$

Return predictability via ML (in practice)

- To determine hyperparameters, one iterates between the training and validation period to choose the hyperparameters that minimize the value of the objective function applied to the resulting forecast errors.
- Each time one re-estimates the model for the new candidate choice of hyperparameters.
- The principle is to simulate a true out-of-sample evaluation of the model to determine the best value for the hyperparameters.
- Roughly speaking, you can think of the joint training and validation period as “in-sample" and the testing period as “out-of-sample" in the terminology seen in the first lectures on return predictability.

Return predictability via ML (in practice)

- As usual, it entails a choice about whether to use fixed, rolling, or expanding (recursive) estimation schemes.
- Gu et al. (2020) apply a hybrid that using a recursive scheme for the training period but a rolling scheme for the validation period.
- This continuously increases the sample size for training the model, yet always tune hyperparameters over a window with a fixed amount of data points, subsequent to the end of the training period.

ML techniques

- We will now (quickly) go through several popular ML techniques that are useful in the context of empirical asset pricing.

Three fundamental elements

Each ML technique may be described by three fundamental elements:

1. The model for $g^*(\cdot)$: Each technique implies some model for describing the general functional form for excess return predictions.
2. The objective function: All techniques employ mean squared prediction error (MSE) as objective function for estimation, yet some with penalty terms used for regularization.
3. Computational algorithms for optimal specification: When one considers non-linear transformations of predictors, their number proliferates and ML techniques entails an efficient approach to search all possible specifications of the model. This point is mostly out of the scope of this course.

Types of (machine) learning

Types of (machine) learning

There exists broadly speaking three types of learning:

1. Supervised learning: Data contains both dependent and independent variables and models are trained to learn the relationship between those, typically used for prediction (example is OLS regression).
2. Unsupervised learning: Data contains only independent variables and models are constructed as to find a structure in this, like grouping or clustering of data (example is PCA).
3. Reinforcement learning: The models learn how to take actions, given data, in an environment through feedback and rewards (example is to play chess).

- Our focus will be mainly supervised learning, yet sometimes with an element of unsupervised learning semi-supervised learning is thus applied.

Simple linear regression model

- The first method cannot really be classified as an ML technique (even though it is actually the starting point of many ML books!)
- This serves as reference point for the performance of the many ML techniques we will consider.

Simple linear model

1. Model: The model is linear in both variables and parameters as

$$g^*(z_{it};\theta) = z_{it}'\theta. \tag{4}$$

2. The objective function: The typical $\mathbb{L}_2$ objective function

$$\mathcal{L}(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(r_{it+1}^e - g(z_{it};\theta))^2. \tag{5}$$

Gu et al. (2020) write about several additional extensions to the objective function which are robust towards heavy tails, among other things.

Penalized linear model

- In the presence of many predictors, the simple linear model is bound to fail since when the number of predictors approaches the number of time series observations, it is both inconsistent and inefficient.
- It will almost surely overfit to noise rather than extracting the signal.
- This is particularly relevant for return predictability where the signal-to-noise ratio is low.

  ➜ A natural solution is to reduce the number of parameters to estimate by sorting out the irrelevant variables known as regularization.

Penalized linear model

1. **Model:** The model is (still) linear in both variables and parameters

$$g^*(z_{it};\theta) = z_{it}'\theta. \tag{6}$$

2. The objective function: The objective function, that we minimize, differs from before by adding a penalty term to the original loss function

$$\tilde{\mathcal{L}}(\theta;\lambda,\alpha) = \mathcal{L}(\theta) + \phi(\lambda,\alpha), \tag{7}$$

where λ,α are non-negative tuning parameters. We will work with the Elastic Net (ENet) formulation as per

$$\phi(\lambda,\alpha) = \lambda(1-\alpha)\sum_{j=1}^{P}|\theta_j| + \frac{1}{2}\lambda\alpha\sum_{j=1}^{P}\theta_j^2, \tag{8}$$

effectively imposing a “budget constraint" on coefficients. In most cases $\alpha = 0.5$ is thought to be defined as ENet, $\alpha = 0$ as LASSO, and $\alpha = 1$ ridge regression.

Penalized linear model

Penalized linear model (cont'd)

- LASSO sets coefficients equal to zero and perform, as such, variable selection.
- Ridge regressions shrinks coefficients to zero, but does not set them exactly to zero.
- ENet is somewhere in between, with the weight given by α.

Dimension reduction: PCR and PLS

- Penalized regressions uses regularization to handle the high-dimensionality of data.
- However, this has problems if the predictors are highly correlated.

Possible issue with regularization

- Suppose, simplistically, we have a (possibly large) set of predictors and suppose they are all equal to the target variable plus iid noise.
- Performing regularization is suboptimal as opposed to taking just the average among all variables and use the average as the predictor.
- Why? Because the averaging cancels out the noise and “extracts" the true signal, yet regularization only removes some of the variables with noise and leaves many that (still) includes noise.

Dimension reduction: PCR and PLS

- This idea with predictor averaging as opposed to predictor regularization/selection is the main idea behind dimension reduction techniques.
- These techniques form linear combinations, for instance the average, of the predictors to help reduce noise and isolate the de-correlated signals in predictors.
- We will see the use of Principal Components Regression (PCR) and Partial Least Squares (PLS).
- Both procedures consist of two steps …

Dimension reduction: PCR and PLS

- …In the first step, PCA combines all regressors into a small set of linear combinations that seeks to extract the common signals among the variables.
- In the second step, those common signals are used in a standard predictive regression (that you have seen in previous lectures).
- That is, PCR regularizes the prediction problem, in some sense, by zeroing out low variance components.
- PCA is inherently unsupervised in the first step, as it does not incorporate the forecast objective when estimating components it condenses information *among* predictors.

Dimension reduction: PCR and PLS

- PLS, on the other hand, conducts the dimension reduction while exploiting the covariance between the target variable and the predictors.
- That is, it essentially constructs components (linear combinations of predictors) that maximise the covariation with the target variable.

PLS components (intuitively)

1. For each predictor j, estimate its univariate return prediction coefficient via OLS. This coefficient reflects the partial sensitivity of returns towards the j'th predictor.
2. Average now all predictors into a single aggregate component with weights proportional to the coefficients from prior step, placing most weight on the most important variable and vice versa.

$\Rightarrow$ as such, PLS performs dimension reduction with the ultimate forecasting objective in mind. To form more than one component, one orthogonalizes the target and the predictors w.r.t. previously constructed components and above procedure is repeated.

Dimension reduction: PCR and PLS

PCR and PLS models

1. Model: In vectorized form the linear regression model is

$$R = Z\theta + E, \tag{9}$$

where R is $NT \times 1$, Z is the $NT \times P$ stacked vector of predictors, and E is a $NT \times 1$ error term (vectorization makes sense since $g(\cdot)$ is independent of i and t). Both PCR and PLS condense all predictors into K predictors, being linear combinations of all P predictors, as per

$$R = (Z\Omega_K)\theta_K + \tilde{E}, \tag{10}$$

where Ω_K is $P \times K$ with columns $\omega_1, \omega_2, \ldots, \omega_K$ that contains the linear combination weights. As such, $Z\Omega_K$ is the dimension-reduced set of predictors.

Dimension reduction: PCR and PLS

PCR and PLS models (cont'd)

2. The objective function: PCR chooses combination weights Ω_K recursively as per

$$\omega_j = \arg max_\omega \text{Var}[Z\omega], \tag{11}$$

subject to $\omega'\omega = 1$ and $\text{Cov}[Z\omega, Z\omega_l] = 0, l = 1,2,\ldots,j-1$. As such, PCR takes the K linear combinations of Z that most faithfully mimic the full predictor set.

PLS seeks the K linear combinations of Z that have maximal predictive association with the forecast target as per

$$\omega_j = \arg max_\omega \text{Cov}[R, Z\omega], \tag{12}$$

subject to $\omega'\omega = 1$ and $\text{Cov}[Z\omega, Z\omega_l] = 0, l = 1,2,\ldots,j-1$.

- Once Ω_K is estimated, the model is estimated by OLS and forecasts generated as in a simple predictive regressions.
- K is the hyperparameter of the procedure.

Generalized linear model

- If the true data generating process is not linear in the variables, but possibly still in the parameters, we may expand the set of predictor by including higher order transforms of them.
- This can be done in numerous ways - here we will consider the most simple way that just computes the square and cubes (or possibly higher orders) of the original predictors.
- However, this proliferates parameters, and requires an additional layer of regularization should be added, here using the ENet as above.
- Define z^p (with abuse of notation) as the vector whose elements are each of z raised to the p'th power.

Generalized linear model

Generalized linear model

1. Model: The model has same format as the penalized linear regression model by including additively the higher-order transformations of the predictors as

$$g(z_{it};\theta) = z_{it}'\theta^{(1)} + z_{it}^{2'}\theta^{(2)}, \ldots, z_{it}^{p'}\theta^{(p)}, \tag{13}$$

   with $\theta = (\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(p)})'$.

2. The objective function: The objective function has same format as the penalized linear model as per

$$\tilde{\mathcal{L}}(\theta; \lambda, \alpha) = \mathcal{L}(\theta) + \phi(\lambda, \alpha), \tag{14}$$

   with the ENet penalty term applied.

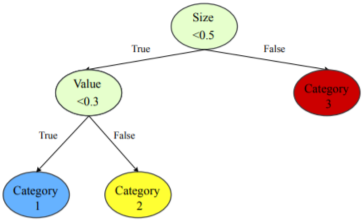
Gu et al. (2020) applies spline function expansions and a related penalty term that are out of the scope here.

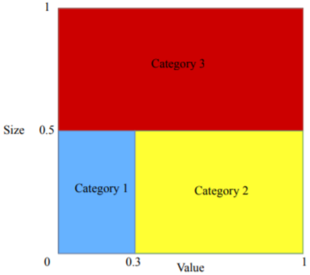
Regression trees

- While the penalized linear model, PCR, and PCA only allow for linear effects of the predictor, the generalized linear model allow for higher-order effects.
- However, there is no way the generalized linear model would be well-specified if we allow it to include interactions as it increases parameters combinatorially.
- As an alternative, regression trees have become a popular ML technique for capturing these types of non-linearities.
- These are fully nonparametric and is fundamentally very different from the linear regression models seen so far.

Regression trees

- Trees are designed to find groups of observations that behave similarly.
- Trees are grown sequentially, where at each step data is sorted into bins based on one of the predictor variables.
- This sequential branching slices the space of predictors into rectangular partitions, which approximates the unknown function $g^*(\cdot)$ with the average value of the outcome variable within each partition.





Note: This figure presents the diagrams of a regression tree (left) and its equivalent representation (right) in the space of two characteristics (size and value). The terminal nodes of the tree are colored in blue, yellow, and red, respectively. Based on their values of these two characteristics, the sample of individual stocks is divided into three categories.

Regression trees

Regression trees

1. Model: The model or prediction of a tree $\mathcal{T}$ with K leaves (terminal nodes - we had three in the example on former slide), and depth L can be written as

$$g(z_{it};\theta,K,L) = \sum_{k=1}^{K} \theta_k 1_{\{z_{it} \in C_k(L)\}}, \tag{15}$$

where $C_k(L)$ is one of the K partitions (or regions) of data.

- Note that each partition is a product of up to L (depth) many indicator functions of the predictions to move down the tree.
- The constant associated with the k'th partition, θ_k, is the sample average of outcomes within the partition.

Regression trees

- In the figure on the former slide, the prediction partition is

$$
\begin{aligned}
g(z_{it};\theta,3,2) &= \theta_1 1_{\{\text{size}_{it}<0.5\}} 1_{\{\text{bm}_{it}<0.3\}} \\
&+ \theta_2 1_{\{\text{size}_{it}<0.5\}} 1_{\{\text{bm}_{it}\geq 0.3\}} \\
&+ \theta_3 1_{\{\text{size}_{it}\geq 0.5\}}
\end{aligned} \tag{16}
$$

- We will not look into the procedures to grow/estimate trees now other than saying they are grown to find the bins that best discriminate among the potential outcomes.
- However, trees are very prone to overfit and some regularization is needed to manage their flexibility.
- It can approximate potentially severe nonlinearities - for instance, a tree of depth L can accommodate up to $(L-1)$-way interactions or very high orders of data!

Regression trees

- We will consider two ways to regularize trees through a concept known as ensemble essentially a combination of models/predictions.

Boosting ensemble learning

- Boosting combines forecasts from many over-simplified trees to form stable and more accurate (less biased) predictions.
- The procedure goes as follows:
  1. Fit a shallow tree (e.g. depth $L = 1$).
  2. Compute residuals and fit those with a second shallow tree of same depth.
  3. Add those two predictions together to form an prediction, but weight/shrink the prediction from the second tree by a factor (known as the learning rate) $\nu \in (0,1)$.
  4. At each new step, fit a shallow tree to the fitted residuals from the model of the prior step, and its predicted residuals are added to the total with a shrinkage ν.
  5. Iterate for B many ensemble trees.
- Tuning parameters are thus L,ν,B.

Regression trees

Random forests

- Random forests (RF) also combines forecasts from many different trees, yet it uses a concept known as bagging.
- The baseline tree bagging procedure draws B different bootstrap samples of the data, fits a separate tree to each, and then average their forecasts.
- Trees for individual bootstrap samples tend to be deep and overfit, yet averaging stabilizes the performance and cancels out the fitting to noise.
- RF use a variant that aims at de-correlating trees to obtain larger benefits from the bagging procedure.

Random forests (cont'd)

- For instance, if firm size is the dominant return predictor in data, most of the bagged trees will have shallow splits on size, resulting in substantial correlation between trees and their forecasts.
- RF de-correlates trees by dropout, which considers only a randomly drawn subset of predictors for splitting at each potential branch.
- This ensures that at least some trees will split ***not*** on size, lowering the correlation among trees and their output, which in turn lowers the variance of the prediction (averages out noise fitting).
- Now the depth L and number of trees B in the bagging are tuning parameters - in fact, so is the number of subset predictors at each split (default is $P/3$ for regressions).

Neural networks

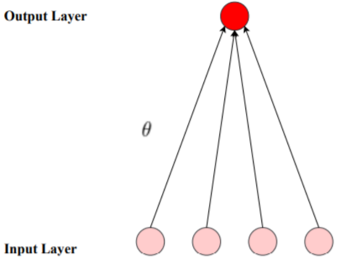
- Another nonlinear prediction method is the artificial neural networks (NN).
- They are typically also known as shallow or deep learning algorithms, being quite complex and typically the least interpretable among ML techniques.
- We focus here, as Gu et al. (2020), on traditional ***feed-forward networks***, which consist of an input layer of raw predictors, one or more hidden layers that interact and nonlinearly transform the predictors, and an output layer that aggregates hidden layers into an ultimate prediction.
- Each layer has a set of neurons, and those layers are connected by synapses hence the name neural network.

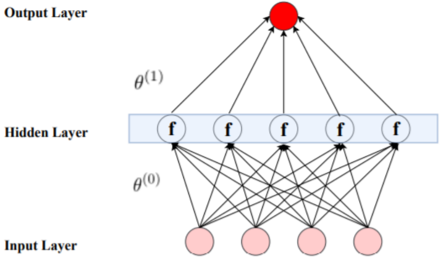
Neural networks

- The number of units in the input layer is equal to the dimension of the predictors (the figure has 4).
- The left panel has no hidden layer and simply aggregates predictors into a prediction via

$$\theta_0 + \sum_{k=1}^{4} \theta_k z_k. \tag{17}$$

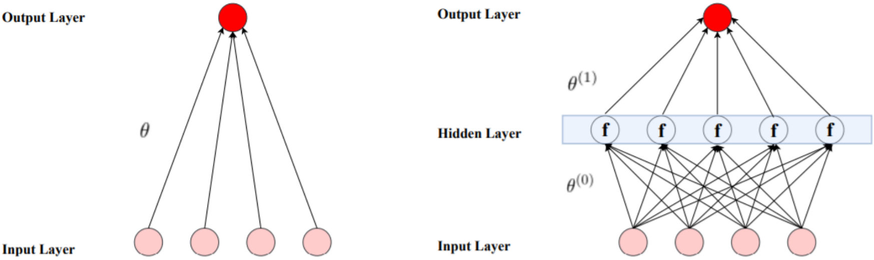
- …which is, indeed, just a linear regression model.





Neural networks

- The right panel has one hidden layer, with each of the five neurons linearly drawing information from all predictors (like the left panel).
- Each neuron applies a nonlinear activation function f that transforms the input before sending its output to the next layer.



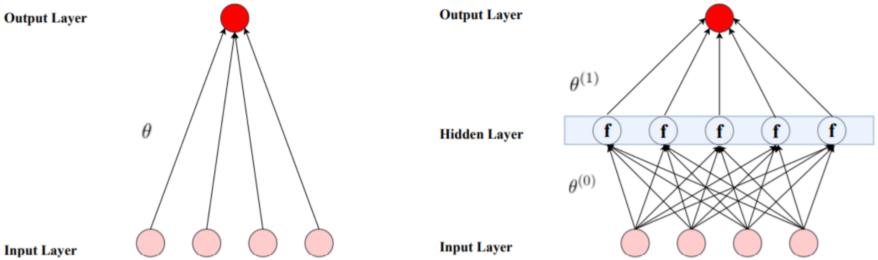
Neural networks

- For instance, the second neuron in the hidden layer transforms inputs into an output as

$$x_2^{(1)} = f\left(\theta_{2,0}^{(0)} + \sum_{k=1}^{4} \theta_{2,k}^{(0)} z_k\right), \tag{18}$$

which in the output layer is aggregated (using input from all neurons) to a final prediction

$$g(z;\theta) = \theta_0^{(1)} + \sum_{j=1}^{5} \theta_j^{(1)} x_j^{(1)}. \tag{19}$$



Neural networks

- In this example, we have $(4+1) \cdot 5 + (5+1) = 31$ parameters with only four predictors, one hidden layer with five neurons!
- An NN that has few (or just one) hidden layers is referred to as shallow, whereas several layers defines deep learning.
- There are **soooo**…many choices to make in designing NNs - Gu et al. (2020) make several and we will not dwell with them here, but you should be aware of those when implementing the method.
- As with the tree-based methods, we will not consider the estimation of NNs, but focus on its implication for the functional form of expected excess returns for now.

Targeting predictors

- You can naturally combine different methods such as targeting which simply introducing a selection step before training a specific model
- Dong et al. (2022) apply ENet before combining forecasts to predict market premium using anomalies (works well and is replicable)
- Borup et al. (2020a) apply ENet before a random forest

Gu, Kelly and Xiu

Overall comments

- We will now go through the paper of Gu et al. (2020)
- The paper is very well-written, and provides a good illustration of how ML can be used to predict returns
- Instead of trying to replicate the findings of Gu et al. (2020), we will take the findings as given focus on how they interpret their results

Data: stock returns

- Monthly CRSP data from 1957 to end of 2016
- The total number of stocks is almost 30,000 and the average number of stocks per month exceeds 6,200.
- The monthly T-bill is used as proxy for the risk-free rate, used to compute excess returns.

Data: predictors

- 94 stock-level predictive characteristics, some of which are updated annually, quarterly, and monthly, obtained from the joint CRSP-Compustat database.
- 74 industry dummies using SIC codes.
- 8 macro-financial predictors common to all stocks from Welch and Goyal (2008) (e.g. dividend-price ratio, term spreads, default spread, and stock market variance) obtained from Goyal's personal website.

$\Rightarrow$ totalling 102 predictors, excluding the 74 industry dummies, without any non-linear effects.

Overarching predictive model

- All the ML techniques we will consider are designed to approximate the overarching model

$$\mathbb{E}_t[r^e_{it+1}|\mathcal{F}_t] = g^*(z_{it}) \tag{20}$$

  defined above.
- Denote the $P_c \times 1$ stock-level predictors by c_{it} for each i, and the $P_x \times 1$ macro-financial predictors by x_t.
- Gu et al. (2020) then consider the total set of, now, stock-level set of predictors as per

$$z_{it} = c_{it} \otimes (1 \;\; x_t), \tag{21}$$

  which is $P_c \cdot (P_x + 1) \times 1$-dimensional and includes all characteristics and their interactions with the common predictors as input.

$\Rightarrow$ totalling $94 \cdot (8+1) = 846$ predictors, excluding the industry dummies.

Overarching predictive model

- To see the relationship to standard β representation models of asset pricing, we may write the conditional model with both time-varying βs and risk premia γs as

$$\mathbb{E}_t[r^e_{it+1}] = \beta_{it}\gamma_t. \tag{22}$$

- Here, stock-level information enters β (which is, indeed, stock-specific) and allows aggregate, common economic information through the dynamic risk premia γ_t.
- Suppose we have K many factors, as usual, and let $\beta_{it} = \theta_1 c_{it}$ and $\gamma_t = \theta_2 x_t$ for some constant parameter matrices $\theta_1(K \times P_c)$ and $\theta_2(K \times P_x)$.

Overarching predictive model

- The β representation of the asset pricing model is the implied as per

$$\begin{aligned} g^*(z_{it}) = \mathbb{E}_t[r^e_{it+1}] = \beta_{it}\gamma_t &= c'_{it}\theta'_1\theta_2 x_t \\ &= (c_{it} \otimes (1 \;\; x_t))'\text{vec}(\theta'_1\theta_2) \\ &\equiv z'_{it}\theta, \end{aligned} \tag{23}$$

  where $\theta = \text{vec}(\theta'_1\theta_2)$.
- As such, the overarching model can be seen within the asset pricing framework.
- Yet, the ML techniques are not restricted to be linear in parameters as used in this example, allowing for a variety of transformations of the z_{it} predictor set.

Sample split

- The total sample period is (initially) split as follows

$$\underbrace{1957M1,\ldots,1974M12}_{\text{training}},\underbrace{1975M1,\ldots,1986M12}_{\text{validation}},\underbrace{1987M1,\ldots,2016M12}_{\text{testing}},$$

using the hybrid scheme between recursively expanding the training period, pushing forward the fixed length validation period to generate forecasts in the testing period.

- Since ML techniques are computationally intensive, they only re-estimate models once a year and not every month.
- The forecast horizon is set to one month or a year.

Performance evaluation

- To assess the predictive ability of the ML techniques across all individual stock excess returns, Gu et al. (2020) use a modified version of the out-of-sample R^2 that you saw in the lectures on return predictability.
- This is defined as

$$R^2_{\text{OS}} = 1 - \frac{\sum_{t=R_2+1}^{T} \sum_{i=1}^{N_{\text{test},t}} (r^e_{it} - \hat{r}^e_{it})^2}{\sum_{t=R_2+1}^{T} \sum_{i=1}^{N_{\text{test},t}} (r^e_{it})^2}. \quad (24)$$

  where $N_{\text{test},t}$ is the number of stocks available at time t.
- This metric is taken only over the testing period.
- Note that the denominator is the sum of squared returns only and does *not* subtract the historical mean, as you have seen in previous lectures.

Performance evaluation

- While subtracting the historical average is common and defines a natural benchmark, Gu et al. (2020) argues that it is flawed for individual stocks.
- They, instead, prefer using zero returns as a natural benchmark.
- The reason is that for individual stocks, the historical average is simply too noisy, being too sensitive to idiosyncratic movements.
- As such, it is so bad that it artificially lowers the bar for predictability.
- All the following results are generally shifted upwards by 3 percentage points when the “usual" R^2_{OS} is applied.

Performance evaluation

- To assess pairwise performance statistically, Gu et al. (2020) use a version of the Diebold-Mariano test.
- It is a cross-sectionally averaged version of the Diebold-Mariano test you have seen in previous lectures.
- The idea is that we construct a loss differential across all stocks under consideration, making a single time series.
- We then form a natural t-statistic upon that.

Performance evaluation

Aggregate Diebold-Mariano test

- To test the forecast performance of method 1 versus method 2, we define

$$
\begin{aligned}
d_{12t} &= \frac{1}{N_{\text{test},t}} \sum_{i=1}^{N_{\text{test},t}} (r_{it}^e - \hat{r}_{1it}^e)^2 - (r_{it}^e - \hat{r}_{2it}^e)^2 \\
&= \frac{1}{N_{\text{test},t}} \sum_{i=1}^{N_{\text{test},t}} \hat{e}_{1it}^2 - \hat{e}_{2it}^2, \qquad (25)
\end{aligned}
$$

  which is the average loss differential across all stocks at time t (within the testing period).

[...]

Performance evaluation

Aggregate Diebold-Mariano test (cont'd)

[...]

- This is, as such, a time series of average loss differentials and we can then test $\mathbb{H}_0 : \mathbb{E}[d_{12t}] = 0$ by running the regression

$$d_{12t} = \psi + \varepsilon_{12t} \tag{26}$$

  and performing a standard t-test on the constant ψ using Newey and West (1987) HAC standard errors to evaluate the null $\psi = 0$.
- This directly takes into account cross-sectional correlation among stocks.

Aggregate Diebold-Mariano test (cont'd)

[...]

- This is equivalent to building the following test statistic

$$t(\bar{d}_{12}) = \frac{\bar{d}_{12}}{\sqrt{\mathrm{Var}[\bar{d}_{12}]}} \xrightarrow{d} N(0,1), \tag{27}$$

where, T_{test} being the length of the testing period,

$$\bar{d}_{12} = \frac{1}{T_{\text{test}}} \sum_{t=R_2+1}^{T} d_{12t}, \tag{28}$$

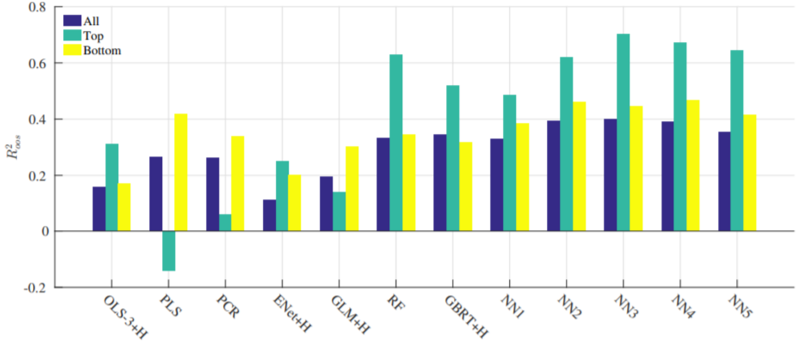
and $\mathrm{Var}[\bar{d}_{12}]$ computed using the classical HAC estimator.

Forecasting Accuracy

Results for the cross-section of individual stocks

Fig. 1: Monthly out-of-sample predictability

	OLS +H	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
All	-3.46	0.16	0.27	0.26	0.11	0.19	0.33	0.34	0.33	0.39	0.40	0.39	0.36
Top 1000	-11.28	0.31	-0.14	0.06	0.25	0.14	0.63	0.52	0.49	0.62	0.70	0.67	0.64
Bottom 1000	-1.30	0.17	0.42	0.34	0.20	0.30	0.35	0.32	0.38	0.46	0.45	0.47	0.42



Note: In this table, we report monthly R^2_{oos} for the entire panel of stocks using OLS with all variables (OLS), OLS using only size, book-to-market, and momentum (OLS-3), PLS, PCR, elastic net (ENet), generalize linear model (GLM), random forest (RF), gradient boosted regression trees (GBRT), and neural networks with one to five layers (NN1–NN5). "+H" indicates the use of Huber loss instead of the l_2 loss. We also report these R^2_{oos} within subsamples that include

Results for the cross-section of individual stocks

Summary of results

- OLS with all predictors are handily dominated by the naive benchmark of zero returns ($R^2_{\mathrm{OS}} < 0$)
- Restricting OLS to a few variables or using penalization (ENet) improves predictability substantially.
- Dimension reduction via PCR or PLS improves predictability further.
- This suggests that single predictors (characteristics) are particularly redundant and fundamentally noisy signals, yet combining then into low-dimensional components averages out noise to better reveal their correlated signals.

Results for the cross-section of individual stocks

Summary of results (cont'd)

- The generalized linear model (that is linear in parameters, but allow for non-linear transformations of variables, yet no interactions) provides no substantial improvement of OLS-3 or ENet.
- Boosted trees and random forests are slightly outperforming PCR and PLS.
- Neural networks are the best performing non-linear method, and the best overall.
- This indicates the value of incorporating complex predictor interactions which are embedded in tree and neural network models, but missed by other techniques.
- Deep learning does not provide much, as NN4 and NN5 fails to improve over NN3.

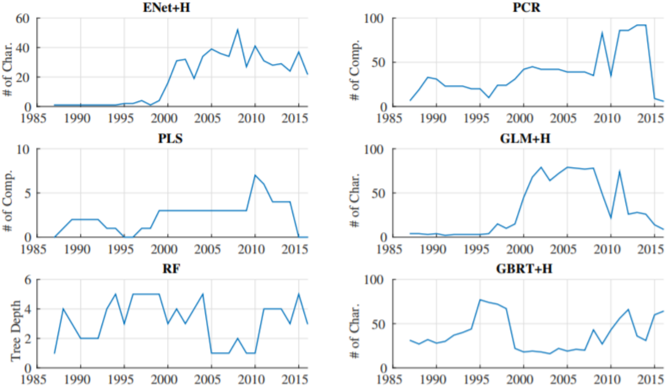
Results for the cross-section of individual stocks

Summary of results (cont'd)

- Separating predictability into that for the 1,000 largest and the 1,000 smallest firms based on market equity allows one to test whether predictability is driven by small, illiquid firms that are poorly priced in the financial market.
- The pattern from above remains in fact, RF and NN generates stronger performance for the largest firms and to a very large degree.
- Since the same pattern persists at annual forecasting horizon, it suggests that ML techniques are able to measure excess returns that persist over business cycle frequencies and not merely short-term inefficiencies.

Results for the cross-section of individual stocks

Fig. 2: Time-varying model complexity



Note: This figure demonstrates the model complexity for elastic net (ENet), PCR, PLS, generalized linear model with group lasso (GLM), random forest (RF) and gradient boosted regression trees (GBRT) in each training sample of our 30-year recursive out-of-sample analysis. For ENet and GLM we report the number of features selected to have non-zero coefficients; for PCR and PLS we report the number of selected components; for RF we report the average tree depth; and for GBRT we report the number of distinct characteristics entering into the trees.

Results for the cross-section of individual stocks

Fig. 3: Pairwise statistical comparison of methods

	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
OLS+H	**3.26***	**3.29***	**3.35***	**3.29***	**3.28***	**3.29***	**3.26***	**3.34***	**3.40***	**3.38***	**3.37***	**3.38***
OLS-3+H		1.42	**1.87**	-0.27	0.62	**1.64**	1.28	1.25	**2.13**	**2.13**	**2.36**	**2.11**
PLS			-0.19	-1.18	-1.47	0.87	0.67	0.63	1.32	1.37	**1.66**	1.08
PCR				-1.10	-1.37	0.85	0.75	0.58	1.17	1.19	1.34	1.00
ENet+H					0.64	**1.90**	1.40	**1.73**	**1.97**	**2.07**	**1.98**	**1.85**
GLM+H						**1.76**	1.22	1.29	**2.28**	**2.17**	**2.68***	**2.37**
RF							0.07	-0.03	0.31	0.37	0.34	0.00
GBRT+H								-0.06	0.16	0.21	0.17	-0.04
NN1									0.56	0.59	0.45	0.04
NN2										0.32	-0.03	-0.88
NN3											-0.32	-0.92
NN4												-1.04

Note: This table reports pairwise Diebold-Mariano test statistics comparing the out-of-sample stock-level prediction performance among thirteen models. Positive numbers indicate the column model outperforms the row model. Bold font indicates the difference is significant at 5% level or better for individual tests, while an asterisk indicates significance at the 5% level for 12-way comparisons via our conservative Bonferroni adjustment.

- To read the table, positive values indicate that the column method outperforms the row method.
- Bold values indicate significance at the 5% significance level.
- Asterisk adjusts for multiple testing via Bonferroni corrections.

Results for the cross-section of individual stocks

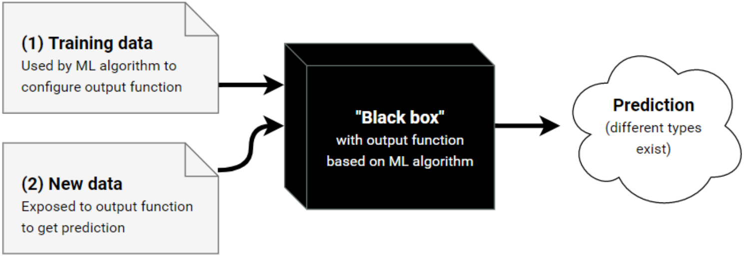
Summary of results

- Constrained linear models - including OLS-3, ENnet, PCR and OLS - produces significant improvements over the unconstrained OLS.
- There is no statistical difference between the penalized OLS model and dimension reduction methods.
- Tree-based methods are marginally (at best) significantly superior to linear models.
- NN are the only models that produce large and significant improvements over linear (and generalized linear) models.
- …yet, they are not significantly better than tree-based methods.

Interpretability

Variable importance

- A very popular critique of ML techniques is that they are simply black box.



- Yet, with better understanding of the algorithms comes more interpretability.

Variable importance

- There are several ways to obtain a sense of the importance of the variables, VI_j used in the ML prediction.
- We consider two sorts for measuring VI_j:
  1. Reduction in R^2_{OS} by setting all values of predictor j to zero, holding the remaining model estimates fixed.
  2. Sum of squared partial derivatives (SSD),

$$SSD_j = \sum_{t=R_2+1}^{T} \sum_{i=1}^{N_{\text{test},t}} \left(\left. \frac{\partial g(z;\theta)}{\partial z_j} \right|_{z_{it}} \right)^2, \tag{29}$$

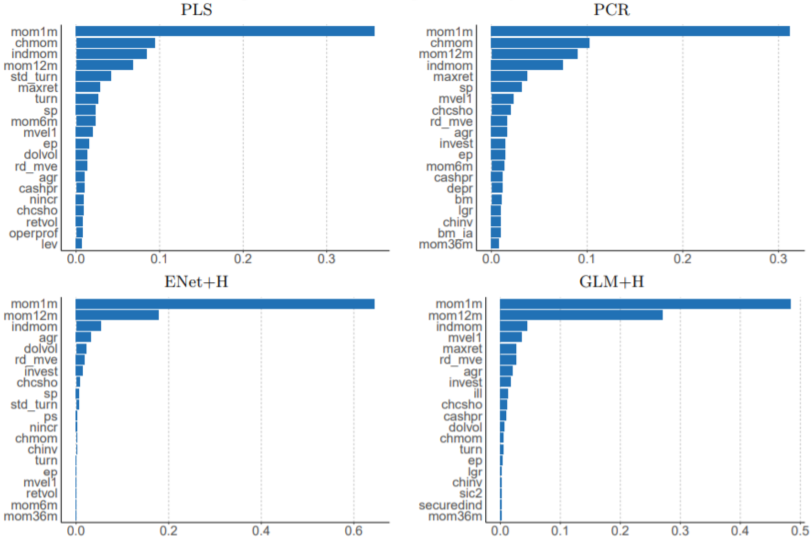
  which summarizes the sensitivity of the model fits to the changes in that variable.
- SSD is not defined for tree-based models (as they are not differentiable) in this case they use the reduction in impurity ($\approx$ MSE).

A note on interpretability

- Here are some very useful references that provides model-agnostic methods:
  1. Interpretable machine learning by Christoph Molnar, link here: `https://christophm.github.io/interpretable-ml-book/`.
  2. Borup et al. (2020c) for application of partial dependency plots and associated variable importance the intuition is quite similar to approaches we will see below by Gu et al. (2020)

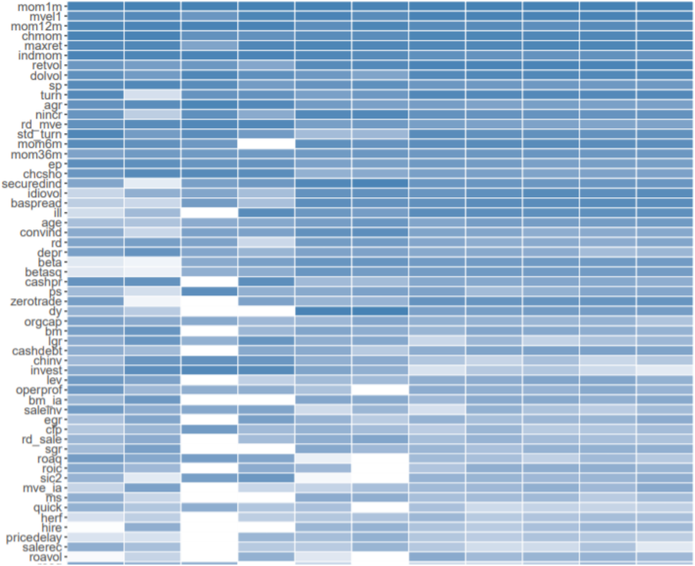
Variable importance

Fig. 4: Variable importance (by R^2_{OS})



Variable importance

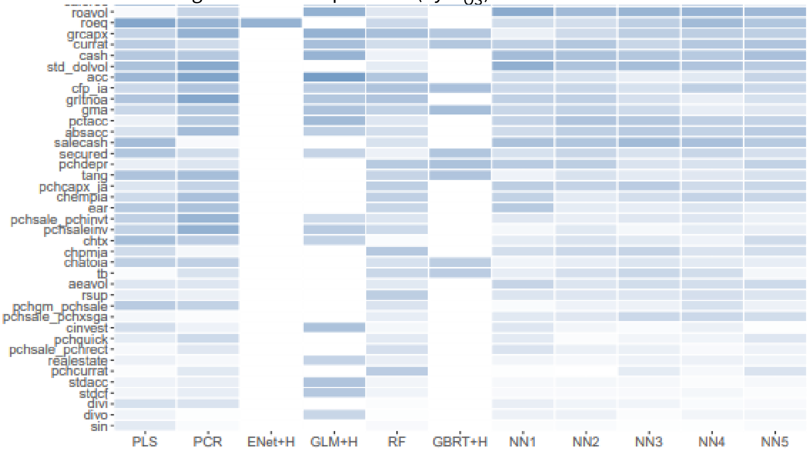
Fig. 5: Variable importance (by R^2_{OS}) across all models



Note: Rankings of 94 stock-level characteristics and the industry dummy (sic2) in terms of overall model contribution

Variable importance

Fig. 6: Variable importance (by R^2_{OS}) across all models



Variable importance

Summary of results

- We may categorize the top predictors into four groups:
  1. Recent price trends: short-term reversal, stock momentum, momentum change, industry momentum, recent maximum return, long-term reversal.
  2. Liquidity variables: turnover, turnover volatility, log market equity, dollar volume, Amihud liquidity, number of zero trading days, bid-ask spread.
  3. Risk measures: total and idiosyncratic volatility, market beta, beta-squared.
  4. Valuation ratios and fundamental signals: earnings-to-price, sales-to-price, asset growth, no. of recent earnings increases.
- Results remain using the SSD measure for variable importance.

Variable importance

- If the top predictors are truly informative, their ranking should be unaffected by the presence of irrelevant variables.

The placebo principle

- To rule out that a given method or procedure mechanically generate the realized results, the placebo principle simulates irrelevant variables that share dynamics with the proposed relevant ones, but they are completely unrelated to the object of interest. Conducting the experiment as with the real data many times using the placebo variables generates an empirical distribution in which one can evaluate the realized values. If the empirical probability of significance is low of the realized values, there is little likelihood that results are spurious and they may be considered robust.
- Use of the placebo principle includes Gu et al. (2020), Borup and Schütte (2021), Borup and Schütte (2020), Delikouras and Kostakis (2019), Kelly and Pruitt (2013) in both cross-sectional asset pricing and (time series) return predictability.

Variable importance

- We can simulate placebo characteristics that have no relation to returns by construction, but have similar properties as the true characteristics as per

$$c_{ijt}^{\text{placebo}} = \frac{2}{N+1}\text{CSrank}(\tilde{c}_{ijt}) - 1, \quad \tilde{c}_{itj} = \rho_j \tilde{c}_{ijt-1} + \epsilon_{ijt},$$

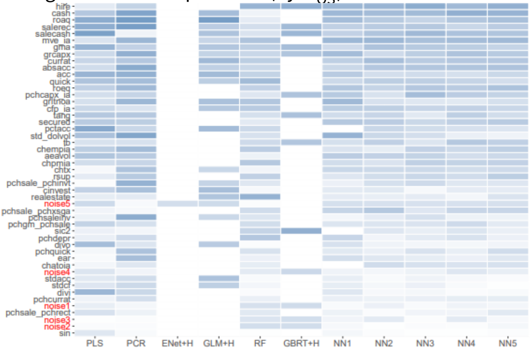
  where $\rho_j \sim U[0.9,1]$, $\tilde{c}$ is the actual value of the characteristic, and CSrank$(\cdot)$ is a cross-sectional rank function such that $c_{ijt}^{\text{placebo}} \in [-1,1]$.
- If we include those placebo variables in the prediction problem as additional predictors and compute the variable importance, we would appreciate them being ranked lowest without affecting the ranking of the remaining variables.

Variable importance (top ones)

Fig. 7: Variable importance (by R^2_{OS}) across all models

Variable importance (bottom ones)

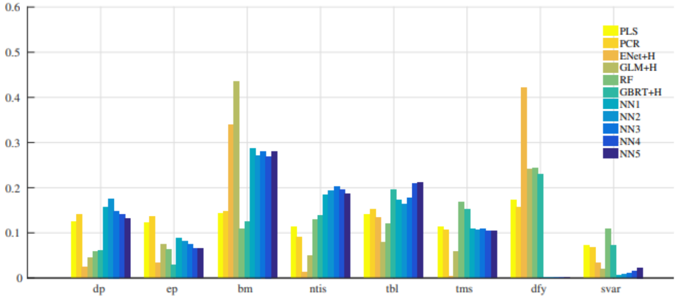
Fig. 8: Variable importance (by R^2_{0S}) across all models



Importance of macro-financial variables

Fig. 9: Variable importance (by R^2_{OS})

	PLS	PCR	ENet+H	GLM+H	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5
dp	12.52	14.12	2.49	4.54	5.80	6.05	15.57	17.58	14.84	13.95	13.15
ep	12.25	13.52	3.27	7.37	6.27	2.85	8.86	8.09	7.34	6.54	6.47
bm	14.21	14.83	33.95	43.46	10.94	12.49	28.57	27.18	27.92	26.95	27.90
ntis	11.25	9.10	1.30	4.89	13.02	13.79	18.37	19.26	20.15	19.59	18.68
tbl	14.02	15.29	13.29	7.90	11.98	19.49	17.18	16.40	17.76	20.99	21.06
tms	11.35	10.66	0.31	5.87	16.81	15.27	10.79	10.59	10.91	10.38	10.33
dfy	17.17	15.68	42.13	24.10	24.37	22.93	0.09	0.06	0.06	0.04	0.12
svar	7.22	6.80	3.26	1.87	10.82	7.13	0.57	0.85	1.02	1.57	2.29



Note: Variable importance for eight macroeconomic variables in each model. Variable importance is an average over all training samples. Variable importances within each model are normalized to sum to one. The lower panel provides a complementary visual comparison of macroeconomic variable importances.

Variable importance

Summary of results

- The book-to-market ratio is a critical predictor across all models.
- Market volatility has little role in any model.
- PCR and PLS puts some weight on all other variables in general, likely because they are highly correlated.
- Linear and generalized linear models strongly favour bond market variables, like the default spread and T-bill.
- Nonlinear models like RF and NN place great emphasis on exactly those ignored by linear models.

Marginal association

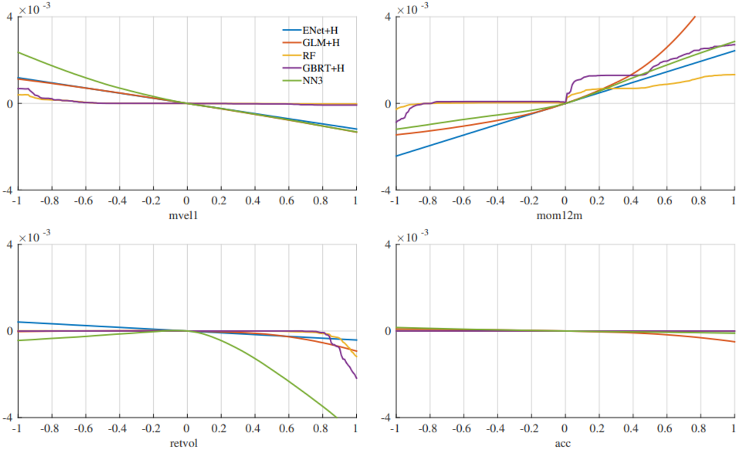
Marginal impact of variable on expected returns

- To trace out the estimated relationship between certain variables and expected returns, Gu et al. (2020) fix all characteristics to their median values except the one variable one is interested in (e.g. size or volatility).
- One then varies the value of the variable under consideration from its minimum to its maximum (for characteristics this is between $[-1,1]$) and compute the fitted value of returns.
- Plotting all fitted values implied from using values in the support of the variables allows one to assess the marginal association.

- This has strong resemblance to partial dependency plots, see e.g. Borup et al. (2020c)

Importance of macro-financial variables

Fig. 10: Marginal association



Note: Sensitivity of expected monthly percentage returns (vertical axis) to individual characteristics (holding all other covariates fixed at their median values).

Marginal association

Summary of results

- ML techniques generally show patterns that are consistent with some well-known phenomena.
- Expected returns are generally decreasing in size, increasing in past one-year returns, and decreasing in volatility.
- Penalized linear regressions (ENet) finds no association between size or volatility with expected returns, strongly opposing nonlinear methods that find large impacts of volatility.

Marginal association

Summary of results (cont'd)

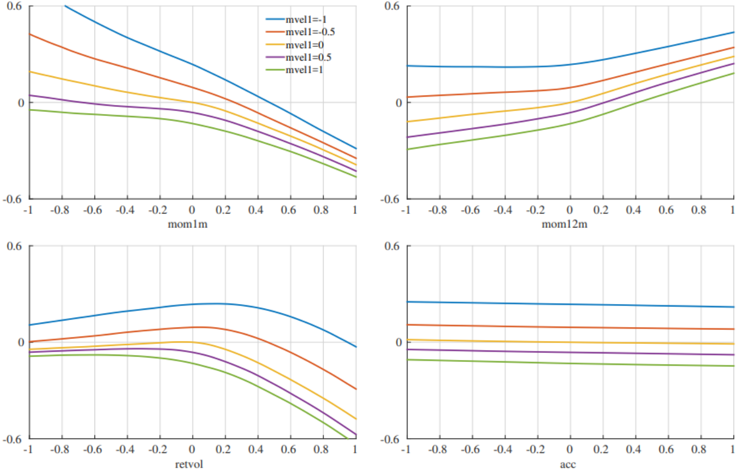
- For example, a firm that drops from the median to the 20th percentile of the size distribution has an increase in expected return (annualized) of approx. 2.4% according to the NN3.
- …and a firm whose volatility increases from the median to the 80th percentile experience an increase in expected return of approx. 3.0% (annualized), according to the NN3.
- Linear models cannot allow for any curvature on these marginal relationships their inability to capture nonlinearities can lead them to prefer a zero association instead, which might explain their differences in performance.

Interaction effects

- The favourable performance of trees and neural networks indicates a benefit to allowing for potentially complex interactions among predictors.
- To understand those effects, we can do as above, yet now we simply vary two variables, simultanously, over their support $[-1,1]$, while still keeping all other variables at their median values.
- Gu et al. (2020) show results for the interaction between size and four variables, namely short-term reversal, momentum, total volatility, and accruals (note that Gu et al. (2020) has a small typo in their text as to which variables the figure includes).
- They also consider the interaction between size and total volatility with aggregated book-to-market ratios and net equity issuance.
- They are all based on NN3.

Marginal association through interactions

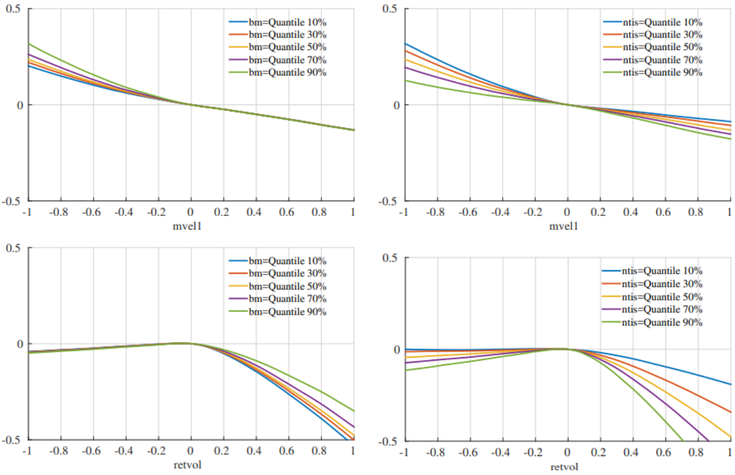
Fig. 11: Marginal association through interactions among characteristics



Note: Sensitivity of expected monthly percentage returns (vertical axis) to interactions effects for mvel1 with mom1m, mom12m, retvol, and acc in model NN3 (holding all other covariates fixed at their median values).

Marginal association through interactions

Fig. 12: Marginal association through interactions among characteristics and macro-financial variable



Note: Sensitivity of expected monthly percentage returns (vertical axis) to interactions effects for mvel1 and retvol with bm and ntis in model NN3 (holding all other covariates fixed at their median values).

Marginal association through interactions

Summary of results

- There are clear interactions among size and short-term reversal, momentum, and total volatility, but none with accruals.
- There are clear interactions among size and aggregate valuations (book-to-market ratios) and equity issuance, and among total volatility and the former two variables.
- Gu et al. (2020) also find that the most important characteristic vs. macro-financial interactions come from interacting a stock's recent price trends (short-term reversal, momentum, industry momentum) with aggregate asset price levels (valuation ratios or T-bill rates).
- …and they are stable over time.

Economic magnitude

Machine learning portfolios

- Rather than assessing pre-specified portfolios, we may also build Machine learning portfolios that seek to explicitly exploit the forecasts.

Machine learning portfolios

1. For each method and at each time point t, gather all one-month forecasts across all stocks.
2. Sort all stocks into deciles based on their forecasted returns (using a univariate portfolio sort)
3. Within each decile portfolio value-weight all stocks.
4. Construct a zero-net investment portfolio that buys the highest expected return stocks (decile 10) and sells the lowest (decile 1).

- Note, that Gu et al. (2020) have not calculated breakpoints using NYSE stocks only...

Machine learning portfolios

Fig. 13: Realized returns in portfolios

	OLS-3+H				PLS				PCR			
	Pred	Avg	Std	SR	Pred	Avg	Std	SR	Pred	Avg	Std	SR
Low(L)	-0.17	0.40	5.90	0.24	-0.83	0.29	5.31	0.19	-0.68	0.03	5.98	0.02
2	0.17	0.58	4.65	0.43	-0.21	0.55	4.96	0.38	-0.11	0.42	5.25	0.28
3	0.35	0.60	4.43	0.47	0.12	0.64	4.63	0.48	0.19	0.53	4.94	0.37
4	0.49	0.71	4.32	0.57	0.38	0.78	4.30	0.63	0.42	0.68	4.64	0.51
5	0.62	0.79	4.57	0.60	0.61	0.77	4.53	0.59	0.62	0.81	4.66	0.60
6	0.75	0.92	5.03	0.63	0.84	0.88	4.78	0.64	0.81	0.81	4.58	0.61
7	0.88	0.85	5.18	0.57	1.06	0.92	4.89	0.65	1.01	0.87	4.72	0.64
8	1.02	0.86	5.29	0.56	1.32	0.92	5.14	0.62	1.23	1.01	4.77	0.73
9	1.21	1.18	5.47	0.75	1.66	1.15	5.24	0.76	1.52	1.20	4.88	0.86
High(H)	1.51	1.34	5.88	0.79	2.25	1.30	5.85	0.77	2.02	1.25	5.60	0.77
H-L	1.67	0.94	5.33	0.61	3.09	1.02	4.88	0.72	2.70	1.22	4.82	0.88
	ENet+H				GLM+H				RF			
	Pred	Avg	Std	SR	Pred	Avg	Std	SR	Pred	Avg	Std	SR
Low(L)	-0.04	0.24	5.44	0.15	-0.47	0.08	5.65	0.05	0.29	-0.09	6.00	-0.05
2	0.27	0.56	4.84	0.40	0.01	0.49	4.80	0.35	0.44	0.38	5.02	0.27
3	0.44	0.53	4.50	0.40	0.29	0.65	4.52	0.50	0.53	0.64	4.70	0.48
4	0.59	0.72	4.11	0.61	0.50	0.72	4.59	0.55	0.60	0.60	4.56	0.46
5	0.73	0.72	4.42	0.57	0.68	0.70	4.55	0.53	0.67	0.57	4.51	0.44
6	0.87	0.85	4.60	0.64	0.84	0.84	4.53	0.65	0.73	0.64	4.54	0.49
7	1.01	0.87	4.75	0.64	1.00	0.86	4.82	0.62	0.80	0.67	4.65	0.50
8	1.16	0.88	5.20	0.59	1.18	0.87	5.18	0.58	0.87	1.00	4.91	0.71
9	1.36	0.80	5.61	0.50	1.40	1.04	5.44	0.66	0.96	1.23	5.59	0.76
High(H)	1.66	0.84	6.76	0.43	1.81	1.14	6.33	0.62	1.12	1.53	7.27	0.73
H-L	1.70	0.60	5.37	0.39	2.27	1.06	4.79	0.76	0.83	1.62	5.75	0.98

Machine learning portfolios

Fig. 14: Realized returns in portfolios

	GBRT+H				NN1				NN2			
	Pred	Avg	Std	SR	Pred	Avg	Std	SR	Pred	Avg	Std	SR
Low(L)	-0.45	0.18	5.60	0.11	-0.38	-0.29	7.02	-0.14	-0.23	-0.54	7.83	-0.24
2	-0.16	0.49	4.93	0.35	0.16	0.41	5.89	0.24	0.21	0.36	6.08	0.20
3	0.02	0.59	4.75	0.43	0.44	0.51	5.07	0.35	0.44	0.65	5.07	0.44
4	0.17	0.63	4.68	0.46	0.64	0.70	4.56	0.53	0.59	0.73	4.53	0.56
5	0.34	0.57	4.70	0.42	0.80	0.77	4.37	0.61	0.72	0.81	4.38	0.64
6	0.46	0.77	4.48	0.59	0.95	0.78	4.39	0.62	0.84	0.84	4.51	0.65
7	0.59	0.52	4.73	0.38	1.11	0.81	4.40	0.64	0.97	0.95	4.61	0.71
8	0.72	0.72	4.92	0.51	1.31	0.75	4.86	0.54	1.13	0.93	5.09	0.63
9	0.88	0.99	5.19	0.66	1.58	0.96	5.22	0.64	1.37	1.04	5.69	0.63
High(H)	1.11	1.17	5.88	0.69	2.19	1.52	6.79	0.77	1.99	1.38	6.98	0.69
H-L	1.56	0.99	4.22	0.81	2.57	1.81	5.34	1.17	2.22	1.92	5.75	1.16

	NN3				NN4				NN5			
	Pred	Avg	Std	SR	Pred	Avg	Std	SR	Pred	Avg	Std	SR
Low(L)	-0.03	-0.43	7.73	-0.19	-0.12	-0.52	7.69	-0.23	-0.23	-0.51	7.69	-0.23
2	0.34	0.30	6.38	0.16	0.30	0.33	6.16	0.19	0.23	0.31	6.10	0.17
3	0.51	0.57	5.27	0.37	0.50	0.42	5.18	0.28	0.45	0.54	5.02	0.37
4	0.63	0.66	4.69	0.49	0.62	0.60	4.51	0.46	0.60	0.67	4.47	0.52
5	0.71	0.69	4.41	0.55	0.72	0.69	4.26	0.56	0.73	0.77	4.32	0.62
6	0.79	0.76	4.46	0.59	0.81	0.84	4.46	0.65	0.85	0.86	4.35	0.68
7	0.88	0.99	4.77	0.72	0.90	0.93	4.56	0.70	0.96	0.88	4.76	0.64
8	1.00	1.09	5.47	0.69	1.03	1.08	5.13	0.73	1.11	0.94	5.17	0.63
9	1.21	1.25	5.94	0.73	1.23	1.26	5.93	0.74	1.34	1.02	6.02	0.58
High(H)	1.83	1.69	7.29	0.80	1.89	1.75	7.51	0.81	1.99	1.46	7.40	0.68
H-L	1.86	2.12	6.13	1.20	2.01	2.26	5.80	1.35	2.22	1.97	5.93	1.15

Note: In this table, we report the performance of prediction-sorted portfolios over the 30-year out-of-sample testing period. All stocks are sorted into deciles based on their predicted returns for the next month. Column "Pred", "Avg", "Std", and "SR" provide the predicted monthly returns for each decile, the average realized monthly returns, their standard deviations, and Sharpe ratios, respectively. All portfolios are value weighted.

What is ML capturing?

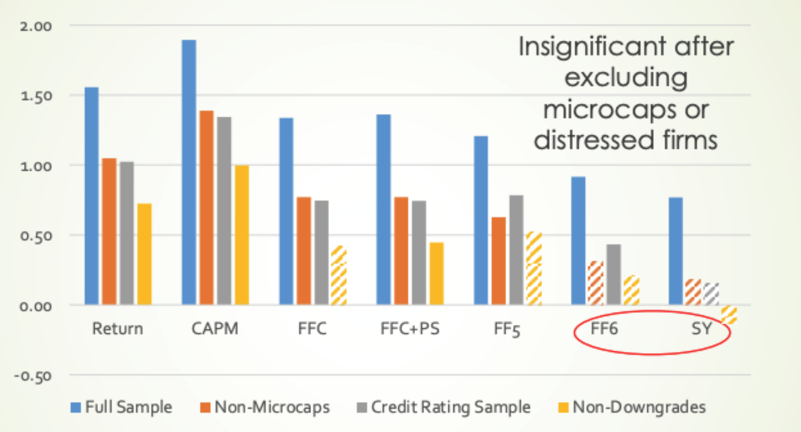
- Avramov et al. (2022) exemines whether machine learning methods clear the common economic restrictions in asset pricing

  ➜ Very similar to your discussion on anomalies!
- Avramov et al. (2022) reexamine the same setup as Gu et al. (2020) for their NN3 model
- Their economic restrictions refers to:

  1. Exclude microcaps: market cap smaller than the 20th NYSE size percentile
  2. Exclude firms with no data on S&P long-term issuer credit rating
  3. Exclude distressed firms: -12 and 12 months around an issuer credit rating downgrade

What is ML capturing?

Fig. 15: After imposing economic restrictions (directly stolen from Avramov et al. (2022))



➜ Returns are roughly halfed when imposing the economic restrictions!

➜ FF6 (FF5 + momentum) α decreases 78% when imposing the restrictions!

Other comments

- Avramov et al. (2022) only compare for linear methods like the IPCA method of Kelly et al. (2019)
- Their evidence suggest that nonlinearities are useful for difficult-to-value/difficult-to-arbitrage stocks!

  ➜ Very similar to our discussion on anomalies!
- Unlike anomalies ML seems to generate returns in the post-2001 period

Other asset classes

- You can find (more or less) a copy-paste of Gu et al. (2020) within other asset classes:
  - Bonds: Bianchi et al. (2021b) or Bianchi et al. (2021a) for the corrected version....
  - Currencies: Filippou et al. (2022)

Examples of ML techniques

- Let us consider numerous examples in the Matlab live script *mlPredictability.mlx*.

Potential projects

About projects

1. Asset return predictability with one or several ML techniques (e.g. on bonds, currencies, equity, etc.)
2. Asset return predictability with other ML type techniques like James-Stein shrinkage, forecast combination (see e.g. Rapach and Zhou (2013), or Borup et al. (2020b)), targeted random forests (Borup et al., 2020a), among other things.
3. Understanding several ML methods' assessment of the variable importance or functional relationship of some important predictors, like the dividend-price ratio.
4. Machine learning portfolios.
5. …or be creative.

$\Rightarrow$ there are several data sources with interesting candidate predictors listed in the document that describes the project.

References

Avramov, D., S. Cheng, and L. Metzker (2022): "Machine learning vs. economic restrictions: Evidence from stock return predictability," *Management Science*.

Bianchi, D., M. Büchner, T. Hoogteijling, and A. Tamoni (2021a): "Corrigendum: Bond risk premiums with machine learning," *The Review of Financial Studies*, 34, 1090–1103.

Bianchi, D., M. Büchner, and A. Tamoni (2021b): "Bond risk premiums with machine learning," *The Review of Financial Studies*, 34, 1046–1089.

Borup, D., B. J. Christensen, N. Mühlbach, and M. S. Nielsen (2020a): "The effects of targeting predictors in a random forest regression model," *Available at SSRN 3551557*.

Borup, D., J. N. Eriksen, M. M. Kjær, and M. Thyrsgaard (2020b): "Predicting bond return predictability," *Available at SSRN*.

Borup, D., D. E. Rapach, E. C. M. Schütte, et al. (2020c): "Now-and backcasting initial claims with high-dimensional daily internet search-volume data," *Available at SSRN 3690832*.

Borup, D. and E. C. M. Schütte (2020): "In search of a job: Forecasting employment growth using Google Trends," *Journal of Business & Economic Statistics*, 1–15.

——— (2021): "Asset pricing with data revisions," *Journal of Financial Markets*, Forthcoming.

Delikouras, S. and A. Kostakis (2019): "A single-factor consumption-based asset pricing model," *Journal of Financial and Quantitative Analysis*, 54, 789–827.

Dong, X., Y. Li, D. E. Rapach, and G. Zhou (2022): "Anomalies and the expected market return," *The Journal of Finance*, 77, 639–681.

Filippou, I., D. Rapach, M. P. Taylor, and G. Zhou (2022): "Out-of-Sample Exchange Rate Prediction: A Machine Learning Perspective," *Available at SSRN 3455713*.

Green, J., J. R. M. Hand, and X. F. Zhang (2017): "Thecharacteristics that provide independent information about average U.S. monthly stock returns," *The Review of Financial Studies*, 30, 4389–4436.

Gu, S., B. Kelly, and D. Xiu (2020): "Empirical asset pricing via Machine Learning," *Review of Financial Studies*, Forthcoming.

Harvey, C. R., Y. Liu, and H. Zhu (2016): "… and the Cross-Section of Expected Returns," *Review of Financial Studies*, 29, 5–68.

Kelly, B. and S. Pruitt (2013): "Market expectations in the cross-section of present values," *The Journal of Finance*, 68, 1721–1756.

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

Masini, R. P., M. C. Medeiros, and E. F. Mendes (2020): "Machine Learning Advances for Time Series Forecasting," *arXiv preprint arXiv:2012.12802*.

McCracken, M. W. and S. Ng (2016): "FRED-MD: A Monthly Database for Macroeconomic Research," *Journal of Business & Economic Statistics*, 34, 574–589.

Mitchell, T. (1997): *Machine Learning*, McGraw Hill.

Newey, W. K. and K. D. West (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

Rapach, D. and G. Zhou (2013): "Forecasting Stock Returns," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier, vol. 2, chap. 6, 328–383.

Samuel, A. L. (1959): "Some studies in machine learning using the game of checkers," *IBM Journal of research and development*, 3, 210–229.

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

Advances in cross-sectional asset pricing

Empirical Asset Pricing

Mads Markvart Kjær

Department of Economics and Business Economics, Aarhus University

E-mail: mads.markvart@econ.au.dk

Spring 2023

Motivation

- Recall that one of the most fundamental research questions in empirical asset pricing is whether a given risk factor is priced in the financial markets and what its compensation is (risk premia).
- We have spend much time on improving the Fama-Macbeth methodology, which is useful in this context.
- Yet, it may still be subject to biases if we omit important *control* factors when estimating and testing a certain risk factor.
- …and/or if the risk factor is poorly measured or generally only weakly related to returns due to noise.
- How can we fix this?

$\Rightarrow$ methods from the ML toolbox, applied in a smart manner, help us!

Omitted variable bias *101*

- To recap the main point of the omitted variable bias concept, let us consider a simple OLS regressions model.

Omitted variable bias (**OVB**)

Suppose the true model (for instance a time-series or cross-sectional stage regression) is

$$y_i = \alpha + \beta x_i + \delta z_i + \varepsilon_i \tag{1}$$

for $i = 1, \ldots, N$ (where i could represent time or assets). Here, x_i and z_i are independent variables that drive the dependent variable y_i, and ε_i is a mean-zero error term. [...]

Omitted variable bias *101*

Omitted variable bias (cont'd)

[...]

- Vectorizing the model reads

$$Y = X\beta + Z\delta + E. \tag{2}$$

- Now suppose we run a regression using only X as regressor, which yields

$$\hat{\beta} = (X'X)^{-1}X'Y. \tag{3}$$

- It can then be shown that

$$\mathbb{E}[\hat{\beta}|X] = \beta + (X'X)^{-1}\mathbb{E}[X'Z|X]\delta. \tag{4}$$

Omitted variable bias *101*

- The omitted variable, z_i introduces a bias unless z_i is unrelated to the variable of interest X
- The bias is a function of the sign of the covariance between x_i and z_i, and the association between z_i and y_i measured by the coefficient δ
- In order to capture the true estimate of β, we are naturallly strongly dependent on specifying a model that properly accounts for all relevant variables...

$\Rightarrow$ this is where ML techniques can be highly useful!

Biases in standard approaches

The model

- To understand the importance of the omitted variable bias and measurement error bias in the context of asset pricing, we consider a simplistic two-factor case like Giglio and Xiu (2021).

The model

- Suppose that $v_t = (v_{1t}, v_{2t})'$ is a vector of two (nontraded) potentially correlated factors.
- They are, without loss of generality, de-meaned.
- Assuming the risk-free rate is observed, and including a constant for testing purposes, we consider the following model

$$r_t^e = \gamma_0 \iota_N + \beta(\gamma + v_t) + u_t, \tag{5}$$

where r_t^e is a vector of the entire universe of excess returns across N assets, and u_t is a mean zero error term.

The model

- Note that the model has the following implication, for the i'th asset

$$\mathbb{E}[r_{it}^e] = \gamma_0 + \beta_i \gamma, \tag{6}$$

which is the standard β representation of expected excess returns, including a constant.

- Our main interest is in estimating the risk premia on (typically a subset of) the factors, which we denote by g_t.
- In this setting, we assume it is $g_t = v_{1t}$, and our goal is to estimate γ_1.

Biases in risk premia estimators

We have in the course examined two (actually three) different ways to test whether a nontraded factor is priced:

Recap: Two approaches to estimating risk premia

- Fama-Macbeth
- Factor-mimicking portfolio

Biases in risk premia estimators

Recap: Fama-Macbeth

- The first approach is to use the two-pass Fama-Macbeth approach.
- Here, risk premia are estimated in the second-pass cross-sectional regression step, using first-pass time series βs as regressors.

- …we are now ready to see how a omitted variable or measurement error bias causes problems in both approaches.

OVB in risk premia estimate

- Let us first consider the implication of omitted variable bias in the Fama-Macbeth setting.
- To make everything clear, note that the model can be expanded to

$$r_t^e = \beta_1(\gamma_1 + v_{1t}) + \beta_2(\gamma_2 + v_{2t}) + u_t, \tag{7}$$

where we in this example assume for simplicity that $\gamma_0 = 0$ for all i.

OVB in risk premia estimate

OVB in Fama-Macbeth

Suppose, possibly guided by a stylistic asset pricing model, that we use only $g_t = v_{1t}$ in our asset pricing analysis. The OVB arises in both the first- and second stage of the Fama-Macbeth procedure:

1. The estimate of β_1 in the first stage is biased as long as the omitted factor v_{2t} is correlated with v_{1t} (and relevant for returns) magnutide of OVB depends on the size of this correlation.
2. The estimate of γ_1 is biased due the former reason *and* that the cross-sectional stage omits $\hat{\beta}_2$ as regressor magnitude of OVB depends on the size of the correlation among β_1 and β_2.

Biases in risk premia estimators

Recap: Factor-mimicking portfolio

- When factors are traded, we can just take its sample mean as estimate of its risk premia.
- This motivates the factor-mimicking approach to estimating risk premia of a nontraded factor by transforming it into a traded portfolio.
- …We construct a traded portfolio by projecting the factor onto a set of traded base asset returns, creating a portfolio that is maximally correlated with g_t.
- The risk premium is then the time series average of this factor-mimicking portfolio return.
- Recall the lecture on portfolio sorting.

OVB in risk premia estimate

OVB in factor-mimicking portfolio

An OVB can arises in the risk premia estimate from the factor-mimicking portfolio if important base assets onto which g_t is project is omitted.

- Suppose our base assets are denoted by $\breve{r}_t^e \subseteq r_t^e$ and we project (regress) our factor $g_t = v_{1t}$ onto those asset together with a constant.
- This yields weights ω^g such that the mimicking portfolio is given by

$$r_t^g = \omega^{g\prime} \breve{r}_t^e. \tag{8}$$

- The expected excess return on this portfolio is

$$\gamma_g^{MP} = \omega^{g\prime} \mathbb{E}[\breve{r}_t^e]. \tag{9}$$

[...]

OVB in risk premia estimate

OVB in factor-mimicking portfolio

[...]

- Since $\breve{r}_t^e \subseteq r_t^e$ it implies that

$$\breve{r}_t^e = \breve{\beta}(\gamma + v_t) + \breve{u}_t. \tag{10}$$

- Using results from OLS regressions, it turns out that

$$\gamma_g^{MP} = \left((\breve{\beta}\Sigma^v \breve{\beta}' + \breve{\Sigma}^u)^{-1} (\breve{\beta}\Sigma^v e_1) \right)' \breve{\beta}\gamma \tag{11}$$

where $e_1 = (1,0)'$ is a column vector with unity in the first element and zero elsewhere, Σ^v is the covariance matrix of the factors, and $\breve{\Sigma}^u$ the covariance matrix of the idiosyncratic risks of the assets used in the projection.

OVB in risk premia estimate

OVB in factor-mimicking portfolio

[...]

- In order to get that $\gamma_g^{MP} = \gamma_1$, we need two conditions:
  1. We need that $\breve{\Sigma}^u = 0$, achieved if, for instance, the bases assets are well-diversified portfolios.
  2. $\breve{\beta}$ is invertible and $v_t = \breve{\beta}^{-1}\breve{r}_t^e$, achieved if the true factors are fully spanned (can be fully recovered by) the base assets.
- In this case, the expression in (11) reduces to the first element in γ, i.e. γ_1.
- If this is not the case, for instance by omitting some important base assets (that causes violation of either of condition 1 and 2 (or, likely, both), we have that the expression does not collapse to γ_1.
- …that is, $\gamma_g^{MP} \neq \gamma_1$, causing a bias in the risk premia estimate.

OVB in risk premia estimate

- The existing literature has typically ignored this bias…
- For instance, most people use a limited set of base assets (e.g. Fama-French type of size-value portfolios).
- Naturally, however, there are other risk sources than size and value which may be correlated with a given factor, for instance consumption growth, which are not captured by those portfolios.
- In this case, the estimator of the risk premia is affected by an OVB.

Measurement error bias in risk premia estimate

- Suppose that the econometrician can only observe the factor subject to some measurement error z_t that is orthogonal to the factors, but possibly correlated with u_t

$$g_t = v_{1t} + z_t. \tag{12}$$

- This is very common for nontraded factors that uses e.g. consumption data or tries to proxy some unobservable metric, like, liquidity.
- This adds another bias to the risk premia estimator in both the Fama-Macbeth and factor-mimicking case.

Measurement error bias in risk premia estimate

- In Fama-Macbeth it causes an attenuation bias in $\hat{\beta}_1$, which in turn leads to a bias in $\hat{\gamma}_1$.
- In the factor-mimicking case, the relevant term expands to

$$\gamma_g^{MP} = \left((\breve{\beta}\Sigma^v\breve{\beta}' + \breve{\Sigma}^u)^{-1}(\breve{\beta}\Sigma^v e_1 + \breve{\Sigma}^{z,u}) \right)' \breve{\beta}\gamma \tag{13}$$

  where $\breve{\Sigma}^{z,u} = \text{Cov}[z_t, u_t]$.
- Unless $\breve{\Sigma}^{z,u} = 0$, measurement errors cause a bias in that $\gamma_g^{MP} \neq \gamma_1$, even under the required conditions above.
- That is, if the measurement error is correlated with idiosyncratic risks of assets, this causes a bias.

A solution

Methodology

- We will now approach a methodology that tackles, jointly, both the omitted variable and measurement error issues.
- Suppose we have p many (possibly unobservable) factors such that

$$\begin{aligned} r_t^e &= \gamma_0 \iota_N + \beta(\gamma + v_t) + u_t, \\ \mathbb{E}[v_t] &= \mathbb{E}[u_t] = 0, \\ \text{Cov}[u_t, v_t] &= 0. \end{aligned}$$

- The objective is to estimate the risk premia of one or more factors g_t without necessarily observing all true factors v_t (because there is no way we can do that!).

Methodology

- To generalize the stylized setting above that assumed $g_t = v_{1t} + z_t$, we now allow for a more broad and general model of the set of d *observable* factors whose risk premia we aim to estimate

$$g_t = \delta + \eta v_t + z_t, \quad \mathbb{E}[z_t] = 0, \; \text{Cov}[z_t, v_t] = 0. \tag{14}$$

- This allows for a slope (η) and intercept (δ) coefficient in the specification, relating the observable factors to all unobservable ones.

Definition of risk premia

- The risk premium of g_t that we want to estimate is the expected excess return of a portfolio with $\beta = 1$ w.r.t g_t and $\beta = 0$ w.r.t. all other factors.
- It turns out that this implies

$$\gamma_g = \eta\gamma. \tag{15}$$

Methodology

- Thus, in order to estimate the risk premia, we need some way to estimate the entire product $\eta\gamma$ or their individual constituents.
- One of the main contributions from Giglio and Xiu (2021) is exactly this identification using an rotation invariance result.

Rotation invariance result (intuitively)

- The rotation invariance result states, intuitively, that the product $\eta\gamma$ can be identified even if one only observes an arbitrary full-rank rotation of the factors.
- That is, if one just observes

$$\hat{v}_t = Hv_t, \tag{16}$$

with H any full-rank $p \times p$ matrix, but does *neither* observe v_t nor H (e.g. negate the expression).

Methodology

- To see this, note that $H^{-1}H = I_p$, such that

$$r_t^e = \gamma_0 \iota_N + \beta H^{-1} H(\gamma - v_t) + u_t \tag{17}$$

and

$$g_t = \delta + \eta H^{-1} H v_t + z_t \tag{18}$$

holds.
- Define now $\hat{\eta} = \eta H^{-1}$, $\hat{\gamma} = H\gamma$, and $\hat{\beta} = \beta H^{-1}$.
- We can now write the model entirely in terms of the rotated factors $\hat{v}_t$.

Methodology

Identify $\hat{\eta}$

- This implies that as long as $\hat{v}_t$ is observed (not the latent v_t) we can identify $\hat{\eta} = \eta H^{-1}$.
- How? Since g_t is observed, and so is $\hat{v}_t$, we can run a regression of g_t onto $\hat{v}_t$, including a constant.
- This estimates δ and $\hat{\eta}$ since the model reads

$$g_t = \delta + \hat{\eta}\hat{v}_t + z_t. \tag{19}$$

Methodology

Identify $\hat{\gamma}$

- This also implies that we can identify $\hat{\gamma} = H\gamma$.
- How? From standard cross-sectional regressions that uses $\hat{\beta} = \beta H^{-1}$ as regressors and average r_t^e as dependent variables a classical Fama-Macbeth with a single cross-sectional regression.
- Those $\hat{\beta}$s can be obtained in a first step from a regression of r_t^e onto $\hat{v}_t$, both being observed.

Methodology

Identify γ_g

- Now, even though we cannot identify η nor γ directly, we can identify their product via above, $\hat{\eta}\hat{\gamma}$, which in turn identifies the risk premia of interest as per

$$\hat{\eta}\hat{\gamma} = \eta H^{-1} H \gamma = \eta\gamma = \gamma_g. \tag{20}$$

- And this trick only applies to γ_g and none of the other quantities in the model.

- Big question is then: *how do we identify these rotated factors $\hat{v}_t$?*

Principal Component Analysis

- …, this is exactly what Principal Component Analysis (PCA) does!
- Recall from past lectures that PCA condenses a large data set into a smaller set of components that aim at capturing the most of the variation in data.
- Ideally, this is used to obtain proxies for the unobserved factors v_t.
- However, the factors are not completely unique in the sense that any rotation of the PCA components would explain the same amount of variance of the original data (for instance multiplying everything by -1).
- That is, PCA really identifies

$$\hat{v}_t = Hv_t \quad not \quad v_t. \tag{21}$$

- …, PCA identifies the factors up to a rotation.

Three-pass methodology

- Using the insights from above, we can formulate the following three-pass procedure in brief terms.

Three-pass procedure

To estimate the risk premia of the (observed) risk factors g_t, denoted γ_g, one conducts the following steps:

1. PCA: Estimate the rotated factors $\hat{v}_t$ via PCA on the full set of returns,
2. $\hat{\gamma}$: Estimate via a standard Fama-Macbeth two-pass cross-sectional regression $\hat{\gamma} = H\gamma$, i.e. the risk premia of the observed $\hat{v}_t$,
3. $\hat{\eta}$: Estimate $\hat{\eta} = \eta H^{-1}$ via a time-series regression of g_t onto $\hat{v}_t$.

The risk premia of g_t, $\eta\gamma$, can then be estimated by taking the product of $\hat{\eta}$ and $\hat{\gamma}$ from steps two and three above.

Three-pass methodology

- To present the methodology more formally, let us formulate the problem in matrix format and ignore the intercept γ_0 for the time being.
- R is the $N \times T$ matrix of all excess returns, V the $p \times T$ matrix of factors, G the $d \times T$ matrix of observable factors, U the $N \times T$ matrix of errors, and Z the $d \times T$ matrix of measurement errors.

$$R = \beta(\gamma \iota_T' + V) + U, \tag{22}$$

  with ι_T a conforming vector of ones.
- Writing $(\bar{R}, \bar{V}, \bar{G}, \bar{U}, \bar{Z})$ as the matrices of de-meaned variables, this rewrites to

$$\bar{R} = \beta\bar{V} + \bar{U}. \tag{23}$$

Three-pass methodology

- The de-meaned version of the observable factors are given by

$$\bar{G} = \eta \bar{V} + \bar{Z}. \tag{24}$$

- To recap, the estimation procedure does *not* require use of the unobservable V, but rather the output from PCA applied to the panel of returns in $\bar{R}$.
- Given observable returns R and factors of interest G, we can write the three steps for the estimator $\gamma_g = \eta\gamma$ more formally in the following slides.

Three-pass methodology

PCA step

- Extract the principal components of returns by conducting the PCA on the $T \times T$ matrix of data $P = (NT)^{-1}\bar{R}'\bar{R}$. This yields the etimated factors as

$$\hat{V} = T^{1/2}(\xi_1,\xi_2,\ldots,\xi_{\hat{p}})'. \tag{25}$$

  where $\xi_1,\xi_2,\ldots,\xi_{\hat{p}}$ are the eigenvectors sorted correspondingly to the largest $\hat{p}$ eigenvalues of P, and $\hat{p}$ is a consistent estimator of the number of factors.
- The estimator of $\hat{p}$ can be based on information criteria or determined by eye balling the plot of eigenvalues/explained variances, or similar arguments. Robustness is important, nevertheless.
- Estimate the factor loadings (risk exposures) as

$$\hat{\beta} = T^{-1}\bar{R}\hat{V}'. \tag{26}$$

Three-pass methodology

Cross-sectional regression step

- Run a standard cross-sectional OLS regression of average returns onto the estimated factor loadings from the prior step to obtain an estimate of the risk premia of the estimated latent factors

$$\hat{\gamma} = (\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}'\bar{r}. \tag{27}$$

- Note that $\bar{r}$ is not the de-meaned returns, but instead their averages, yet otherwise of same dimension.
- FYI, it turns out OLS is actually the efficient estimator here.

Three-pass methodology

Time-series regression step

- Run a standard time-series OLS regression of g_t onto the extracted factors from step 1 and obtain the slope estimator and the fitted value of the observable factor denoted by $\hat{G}$ as

$$\hat{\eta} = \bar{G}\hat{V}'(\hat{V}\hat{V}')^{-1}, \quad \text{and} \quad \hat{G} = \hat{\eta}\hat{V}. \tag{28}$$

Risk premia estimate

- Obtain the risk premia estimate as

$$\hat{\gamma}_g = \hat{\eta}\hat{\gamma}. \tag{29}$$

- This can also be expressed compactly by combining some of the components above

$$\hat{\gamma}_g = \bar{G}\hat{V}'(\hat{V}\hat{V}')^{-1}(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}'\bar{r}. \tag{30}$$

Three-pass methodology

- The estimator can allow for the estimation of the intercept as well, writing excess returns as

$$r_t^e = \gamma_0 \iota_N + \beta(\gamma + v_t) + u_t. \tag{31}$$

- This simply amounts to altering step 2 (cross-sectional step) to include a constant, leading to the following compact forms

$$\begin{aligned} \hat{\gamma}_0 &= (\iota_N' M_{\hat{\beta}} \iota_N)^{-1} \iota_N M_{\hat{\beta}} \bar{r} \\ \hat{\gamma}_g &= \bar{G} \hat{V}' (\hat{V} \hat{V}')^{-1} (\hat{\beta}' M_{\iota_N} \hat{\beta})^{-1} \hat{\beta}' M_{\iota_N} \bar{r}. \end{aligned} \tag{32}$$

where

$$\begin{aligned} M_{\hat{\beta}} &= I_N - \hat{\beta}(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}' \\ M_{\iota_N} &= I_N \iota_N (\iota_N' \iota_N)^{-1} \iota_N'. \end{aligned} \tag{33}$$

- Note that this require very little extra work - it just utilizes everything obtained in the steps without a constant.

Three-pass methodology

- Essentially the third step is the new bit compared to a standard Fama-Macbeth analysis that uses statistical factors (from step 1) in addition to the observable ones.
- This last step is critical, because it translates the uninterpretable risk premia of the latent factors to those the economic theory predicts.
- Also, it removes measurement error effects since we use only the effect in $\hat{\eta}$ and not the entire G in our procedures.
- $\ldots$, in fact $\hat{G} = \hat{\eta}\hat{V}$ is the factors cleaned from measurement error.
- Why? Note that

$$\hat{G} = \hat{\eta}\hat{V} = \bar{G} - \hat{Z}, \tag{34}$$

so that $\hat{G}$ is $\bar{G}$ when the (estimated) measurement error has been removed.

Three-pass methodology

Separation principle

- Even though g_t can be multi-dimensional ($d > 1$), the estimation of risk premia for each observable factor is separate (simply because we control for everything via $\hat{V}$).
- This is in contrast to standard Fama-Macbeth that ***has*** to include all relevant factors in the second-stage cross-sectional regression.

Interpretation

Interpretations

- This three-pass procedure is an extension of both the two-pass Fama-Macbeth procedure and the factor-mimicking approach.
- These two typically give very different results - now they do *not*.
- How to interprete $\hat{\eta}$ differs, however, a bit depending on whether we focus on the Fama-Macbeth or the factor-mimicking interpretation...

FM interpretation

- The invariance result supports directly a Fama-Macbeth interpretation
- $\hat{\eta}$ tells us how to rotate the estimated model such that $\hat{g}_t$ appears as the first factor
- $\hat{g}_t$ is used as the observables factor(s), together with $p - d$ many PCs as controls.
- This ensures no OVB in both the first and second stage of the Fama-Macbeth procedure.

Factor-mimicking portfolio interpretation

- Is also lends support to a factor-mimicking interpretation where the factor(s) is projected onto the PCs which constitute the base assets.
- Why? Because they exactly average out noise (remember the motivation from past lectures of PCA) such that $\Sigma^u \approx 0$ and they span the returns by construction of the factors.
- $\dots$, i.e. it satisfies the conditions for no OVB in the factor-mimicking portfolio.
- Using this interpretation, $\hat{\eta}$ denotes the factor weights $\rightarrow$ with risk premium of the weights times the RP of the portfolios $\hat{\gamma}$

Asymptotic theory

Theorem: Consistency

Under appropriate assumptions, and if $\hat{p} \xrightarrow{p} p$, then as $N, T \to \infty$,

$$\hat{\eta} \xrightarrow{p} \eta H^{-1}, \; \hat{\gamma} \xrightarrow{p} H\gamma, \; \text{and} \; \hat{\eta}\hat{\gamma} = \hat{\gamma}_g \xrightarrow{p} \gamma_g. \tag{35}$$

for some invertible (with probability one) matrix H.

- This implies that the three-pass procedure indeed produces the right estimates, as long as we consistently estimate the correct number of latent factors.

Asymptotic theory

Theorem: Asymptotic normality

Under appropriate assumptions, and if $\hat{p} \xrightarrow{p} p$, then as $N, T \to \infty$, together with $T^{1/2}N^{-1} \to 0$, we have that

$$T^{1/2}\hat{\gamma}_g \xrightarrow{d} N(\gamma_g, \Phi), \tag{36}$$

where the covariance matrix is given by

$$\begin{aligned}\Phi = &(\gamma'(\Sigma^v)^{-1} \otimes I_d)\Pi_{11}((\Sigma^v)^{-1}\gamma \otimes I_d) \\ &+ (\gamma'(\Sigma^v)^{-1} \otimes I_d)\Pi_{12}\eta' + \eta\Pi_{21}((\Sigma^v)^{-1}\gamma \otimes I_d) + \eta\Pi_{22}\eta',\end{aligned}$$

where Π is the covariance matrix of $z_t\hat{v}_t$.

Asymptotic theory

- Giglio and Xiu (2021) show that putting “hats" on everything in Φ to get $\hat{\Phi}$ yields $\hat{\Phi} \xrightarrow{p} \Phi$, provided that we use HAC type estimators.
- Under similar assumptions above, we can also get asymptotic normality of both $\hat{\gamma}_0$ and $\hat{\gamma}_g$ when we include a constant in the model.
- The variance expressions are much more involved (see Online Appendix to Giglio and Xiu (2021)). though already implemented in the live script provided to you.
- It is also quite interesting that the variance estimator does not require any EIV adjustment, like e.g. in Shanken (1992).
- And the variance estimator acts asymptotically as if all factors were observable - even though we have to estimate (a rotation of) them in step 1 via PCA.

Hypothesis testing

Hypothesis testing

With those results at hand, one can construct standard hypothesis tests on the significance of the risk premia as follows

$$t(\hat{\gamma}_g) = \frac{\hat{\gamma}_g}{\sqrt{\mathrm{Var}[\hat{\gamma}_g]}} \xrightarrow{d} N(0,1). \tag{37}$$

Ridge regression

- Remember from previous lectures that PCA essentially functions as a regularization technique that weights the relevant elements in a given data set according to their contribution to the signal.
- Another type of regularization is a version of the ENet known as ridge regression, which was defined by $\alpha = 1$ in the penalty term $\phi(\lambda, \alpha)$ in the lectures on ML and return predictability:

$$\tilde{\mathcal{L}} = \mathcal{L} + \frac{1}{2}\lambda \sum_{j=1}^{P} \theta_j^2 \tag{38}$$

- In the present context, one could imagine that instead of using PCA to select the relevant factors and project a given factor onto those for obtaining the factor-mimicking weights, one could instead use ridge regression.
- ..., that is, use ridge regression to conduct the (otherwise infeasible OLS) projection of the factor onto all available returns, handling parameter proliferation.

Ridge regression

- Giglio and Xiu (2021) shows the resulting risk premia estimator, discarding it in the same vein, will not be as efficient as the PCA approach.
- Why? For exactly the same reason we motivated dimension reduction techniques in previous lectures.
- …the ridge puts weight on all assets and will reflect some of the noise inherent in the assets.
- …PCA averages out the noise by forming the optimal linear combinations that extract the relevant signal.

Testing factor strength

Testing the strength of an observed factor

- A general and important concept in empirical asset pricing is so-called weak factors.

Weak factors

Weak factors are observable factors that are only weakly reflected in the cross-section of test assets.

- These typically challenge econometric techniques and the economic story.
- We can use the present framework to test the strength of a given observable factor.

Testing the strength of an observed factor

- Note that in the model

$$g_t = \delta + \eta v_t + z_t \tag{39}$$

  an $\eta \approx 0$ indicates that either measurement error dominates or the factor is not pervasive, i.e. not a strong factor.
- In those cases, factor loadings/risk exposures would be weak (poorly estimated with little cross-sectional variation).

Testing strength (intuitively)

- The R^2 from the regression of g_t onto the estimated latent factors $\hat{v}_t$ measure the strength of the factor.
- The significance of this strength can be tested via $\mathbb{H}_0 : \eta = 0$ vs. $\mathbb{H}_1 : \eta \neq 0$.
- Those hypotheses are implied by similar hypotheses on the observed/estimated $\hat{\eta}$.

Testing the strength of an observed factor

Testing strength

To test $\mathbb{H}_0 : \eta = 0$ vs. $\mathbb{H}_1 : \eta \neq 0$ we may formulate a conventional Wald test as per

$$\hat{W} = T\hat{\eta}\left((\hat{\Sigma}^v)^{-1}\hat{\Pi}_{11}(\hat{\Sigma}^v)^{-1}\right)^{-1}\hat{\eta}' \xrightarrow{d} \chi^2_{\hat{p}}. \tag{40}$$

Moreover, as measure of strength we can use the (estimated) R^2 from the same regression as

$$\hat{R}^2_g = \frac{\hat{\eta}\hat{V}\hat{V}'\hat{\eta}'}{\bar{G}\bar{G}'} \xrightarrow{p} R^2_g \tag{41}$$

where R^2_g is the true version from a regression using v_t, not $\hat{v}_t$.

- This naturally hinges upon the assumption that the latent factors are pervasive/strong otherwise $\eta \neq 0$ wouldn't mean anything.

Testing the strength of an observed factor

- A few remarks:
  1. The degrees of freedom in the Wald test is $\hat{p}$ because we test this many null restrictions in η.
  2. This is also why we do a Wald test and not a standard t-test which would only apply if we use one latent factor $\hat{p} = 1$.
  3. We say a factor is strong (with strength $\hat{R}^2_g$) if we reject $\mathbb{H}_0 : \eta = 0$, i.e. if $\hat{W}$ exceeds the $\chi^2_{\hat{p}}(1-\alpha)$ percentile for α significance level.
- Giglio and Xiu (2021) show that

$$\mathbb{P}\left(\hat{W} > \chi^2_{\hat{p}}(1-\alpha)|\mathbb{H}_0\right) \to \alpha \text{ and } \mathbb{P}\left(\hat{W} > \chi^2_{\hat{p}}(1-\alpha)|\mathbb{H}_1\right) \to 1.$$

as $N, T \to \infty$ (size and consistency of the test).

Empirical illustration

The objective

- Instead of simply go through the results of Giglio and Xiu (2021), we will instead use the same portfolios but consider a different sample period
- You should note that Giglio and Xiu (2021) do not include an intercept in their main table
- A small hint: You should, furthermore, try to examine whether the results are robust towards including more portfolios than the 202...
- ... but now for our object:

The objective of the empirical study

We will estimate the risk premia on some traded and nontraded factors and compare their values and conclusions to the conventional approaches of Fama-Macbeth

- ... I leave the comparison with the factor-mimicking portfolio as an exercise for you...

Data on test assets

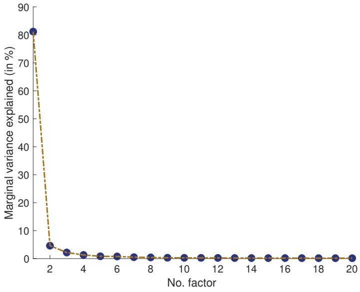
- Giglio and Xiu (2021) consider many different asset classes. We will, however, focus on the US equities from (July) 1963 to end of 2020
- We will consider 202 standard equity portfolios (pfs) from Kenneth French's website which span the most well-known dimensions of risk:
  1. 25 pfs sorted on size and book-to-market ratio
  2. 17 industry pfs
  3. 25 pfs sorted on profitability and investment
  4. 25 pfs sorted on size and variance
  5. 35 pfs sorted on size and net issuance (Note that this does not exists... We have included "Net Share Issues". These 35 portfolios have the same ticker as the one from the code of Giglio and Xiu (2021). Furthermore, the Kenneth French website states this is a 5x5 sort but the file includes 35 portfolios. I have no idea on who is right...)
  6. 25 pfs sorted on size and accruals
  7. 25 pfs sorted on size and beta
  8. 25 pfs sorted on size and momentum

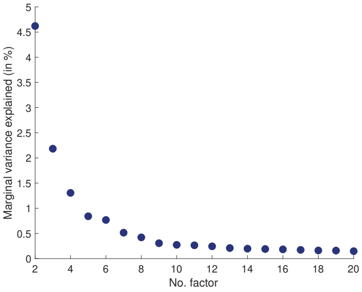
Data on observable factors g_t

- They focus on a selected few factors that have been discussed in the asset pricing literature.
- They consider single risk factors or groups of those this has no influence on the three-pass estimator whatsoever, but it influences the Fama-Macbeth estimator quite some.
- They consider the market excess return, SMB, HML, profitability (RMW), investment (CMA), momentum (MOM), betting-against-beta (BAB), quality (QMJ), and AR(1) innovations in industrial production grwoth, VAR(1) innovations in the first three PCs of 279 macro-finance variables, liquidity, two intermediary capital factors, high temperatures in Manhatten, global land surface temperatures, quasiperiodic Pacific Ocean temperatures, and the number of sunspots, and consumption growth.
- See the Online Appendix in Giglio and Xiu (2021) for data sources.
- We will, instead examine the CAPM, FF3 and innovations in Industrial product.

No. of factors from $\bar{R}$

- In order to estimate the number of factors $\hat{p}$, we may examine their individual variance contribution (or, equivalently, eigenvalues).





No. of factors from $\bar{R}$

- The first factor contributes by far the most, and the marginal variance explained generally show a hockey stick format.
- Zooming in, shows a drop after the 4th factor
- For the interested student, Giglio and Xiu (2021) also provide a concise estimator in their Online Appendix for the number of factors this also suggest 4 factors in our case

  $\rightarrow$ so let us go with 4, but note that this is our choice.
- We generally highlight the need for robustness, and a rule of thumb is to rather choose a factor too many than too few (such that we do not have an OVB issue).

No. of factors from $\bar{R}$

- We can receive a more quantitative measure to motivate our choice by looking at the R^2 of the factor model when increasing the number of included factors:

	1	2	3	4	5	6	7	8	9	10
R^2	0.00	0.17	0.28	0.64	0.64	0.64	0.66	0.67	0.68	0.68

Risk premia estimates

- Let's consider some results - We will focus on a test for the CAPM and innovations in Industrial production
- Let us start with the risk premia estimates of the two-pass and three-pass method (the estimates are measured in %)

	Avg return	two-pass	three-pass
Market	0.57	-0.10	0.37
		[-0.36]	[1.87]
ΔIP	(-)	22.56	0.64
		[1.48]	[0.55]

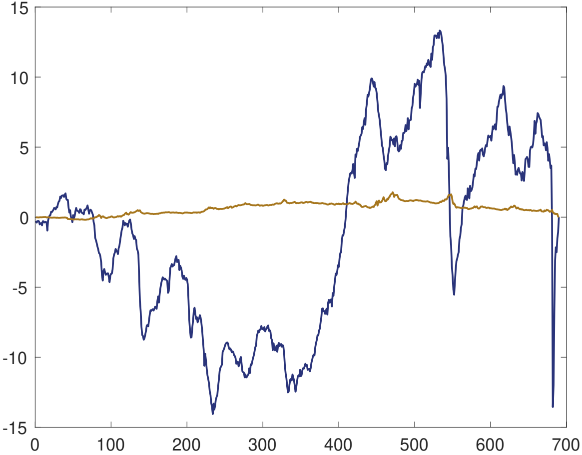
- extremely large difference in the estimates of the two-pass and three-pass!

Test for strength of the factors

- Next, we can examine whether the factors are strong:

	factor weak (p-value)	R_g^2
Market	(0.00)	0.99
ΔIP	(0.93)	0.00

A weak factor plot



Risk premia estimates

The empirical results of Giglio and Xiu (2021) is hidden in the appendix B1. Some of the conclusions are listed down below:

A small note to the results in Giglio and Xiu

1. The Fama-Macbeth estimator is subject to i) OVB as risk premia estimates vary substantially by controls (and differ from averages of traded factors) and ii) measurement error bias by finding significant risk premia on e.g. pure noise factors related to the climate, which the three-pass methodology deems as dominated by noise.
2. The three-pass methodology delivers risk premia estimates that are close economically and statistically to the averages of traded factors, and finds that several of the nontraded factors are significantly priced.
3. The strength (measured by R_g^2 and test via $\hat{W}$) indicate that traded factors are generally strong, while many nontraded ones may be deemed weak.

References

Feng, G., S. Giglio, and D. Xiu (2020): "Taming the Factor Zoo: a test of new factors," *The Journal of Finance*, Forthcoming.

Giglio, S. and D. Xiu (2021): "Asset pricing with omitted factors," *Journal of Political Economy*, Forthcoming.

Shanken, J. (1992): "On the estimation of beta-pricing models," *The Review of Financial Studies*, 5, 1–33.

Advances in cross-sectional asset pricing: Taming the Factor Zoo

Empirical Asset Pricing

Mads Markvart Kjær

Department of Economics and Business Economics, Aarhus University

E-mail: mads.markvart@econ.au.dk

Spring 2023

A zoo of factors

As we went through a couple of weeks ago

- The finance literature has constructed an enormous amount of factors that apparently all are priced in the cross-section of U.S. stocks
- A substantial part of the analysis is to evaluate the new proposed risk factor by existing risk factors
- A typical saying is something like: "The findings cannot be explained by existing risk factors"
- ... or "The findings cannot be explained by standard risk factors"
- However, as we will see; this is done in a bit ad-hoc manner traditionally...

The standard approach of evaluating risk factors

The Review of Financial Studies

Short-term Momentum

Mamdouh Medhat
Cass Business School, City, University of London

Maik Schmeling
Goethe University Frankfurt and Centre for Economic Policy Research (CEPR)

significant average return of +16.4% per annum (Figure 1, right bar). We show that both strategies generate significant abnormal returns relative to the standard factor models currently applied in the literature. We also show that short-term momentum persists for 12 months and is strongest among the largest and most liquid stocks. Finally, we show that our main findings extend to 22 developed markets outside the United States.

The standard approach of evaluating risk factors

Journal of Financial Economics 135 (2020) 725–753

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

ELSEVIER

Left-tail momentum: Underreaction to bad news, costly arbitrage and equity returns ☆

Yigit Atilgan[a], Turan G. Bali[b,*], K. Ozgur Demirtas[a], A. Doruk Gunaydin[a]

[a] Sabanci University, School of Management, Orhanli Tuzla 34956, Istanbul, Turkey
[b] Georgetown University, McDonough School of Business, Washington, D.C. 20057, USA

Table 2
Univariate portfolio analysis.

This table presents return comparisons between equity deciles formed monthly based on VaR1 between 1962 and 2014. Portfolio 1 is the portfolio of stocks with the lowest value-at-risk and Portfolio 10 is the portfolio of stocks with the highest value-at-risk. The table reports the one-month-ahead excess returns and five-factor alphas for each decile. The last column shows the differences of monthly excess returns and alphas between deciles 10 and 1. Alphas are calculated after adjusting for the market, size, value, and momentum factors of Fama and French (1993) and Carhart (1997) and the liquidity factor of Pastor and Stambaugh (2003). Panel A presents results for value-weighted portfolio returns. Panel B presents results for equal-weighted portfolio returns. VaR1 is defined in Table 1. Newey-West (1987) adjusted *t*-statistics are presented in parentheses.

Panel A: Value-weighted returns	Port1	Port2	Port3	Port4	Port5	Port6	Port7	Port8	Port9	Port10	High-Low
Excess return	0.47	0.61	0.58	0.56	0.57	0.57	0.62	0.51	0.31	−0.31	−0.78
	(3.28)	(3.59)	(3.14)	(2.72)	(2.49)	(2.20)	(2.25)	(1.70)	(0.90)	(−0.77)	(−2.34)
Alpha	0.07	0.09	0.02	−0.07	−0.05	−0.08	−0.01	−0.16	−0.39	−0.87	−0.94
	(0.92)	(1.62)	(0.35)	(−0.99)	(−0.56)	(−0.88)	(−0.12)	(−1.34)	(−3.16)	(−5.02)	(−4.42)

Panel B: Equal-weighted returns	Port1	Port2	Port3	Port4	Port5	Port6	Port7	Port8	Port9	Port10	High-Low
Excess return	0.71	0.81	0.85	0.88	0.92	0.89	0.88	0.68	0.52	0.05	−0.66
	(4.73)	(4.65)	(4.42)	(4.20)	(4.01)	(3.59)	(3.37)	(2.36)	(1.64)	(0.15)	(−2.34)
Alpha	0.25	0.25	0.21	0.19	0.19	0.13	0.11	−0.09	−0.22	−0.56	−0.80
	(2.99)	(3.63)	(3.19)	(3.11)	(3.07)	(1.99)	(1.93)	(−1.36)	(−2.85)	(−5.50)	(−5.20)

The standard approach of evaluating risk factors

A Market-Based Funding Liquidity Measure

Zhuo Chen
PBC School of Finance, Tsinghua University

Andrea Lu
Faculty of Business and Economics, The University of Melbourne

We construct a traded funding liquidity measure from stock returns. Guided by a model, we extract the measure as the return spread between two beta-neutral portfolios constructed using stocks with high and low margins, to control for their sensitivity to the aggregate funding shocks. Our measure of funding liquidity is correlated with other funding liquidity proxies. It delivers a positive risk premium that cannot be explained by existing risk factors. A model augmented by our funding liquidity measure has superior pricing performance for various portfolios. Despite evident comovement, this measure contains additional information that is not subsumed by market liquidity. (*JEL* G10, G11, G23)

Received March 29, 2017; accepted August 8, 2018 by Editor Wayne Ferson.

The standard approach of evaluating risk factors

	(0.22)	(1.28)	(0.5)	(1.29)	(2.28)	(2.27)
F. Average across five margin-sorted portfolios [1/1965–10/2012][a]						
Exret	0.32	0.55	0.52	0.73	1.21	0.90
	(2.30)	(3.87)	(3.73)	(4.98)	(6.97)	(5.77)
Alpha	0.12	0.25	0.17	0.34	0.76	0.64
	(0.97)	(1.92)	(1.31)	(2.38)	(3.31)	(2.93)

This table presents BAB portfolio returns conditional on the five margin proxies and the average portfolio returns across five margin proxies. Size refers to a stock's market capitalization. Idiosyncratic volatility is calculated following Ang et al. (2006). The Amihud illiquidity measure is calculated following Amihud (2002). Institutional ownership refers to the fraction of common shares held by institutional investors. Analyst coverage is the number of analysts following a stock. Stocks are sorted into five groups based on NYSE breaks, where 1 indicates the low-margin group and 5 indicates the high-margin group. The high-margin group includes stocks that have small market cap, large idiosyncratic volatility, low market liquidity, low institutional ownership, and low analyst coverage. "Diff" indicates the return difference between two BAB portfolios constructed with high-margin and low-margin stocks. We report raw excess returns (indicated by "Exret") and risk-adjusted alphas. Alphas are calculated using a five-factor model: the Fama-French (1993) three factors, the Carhart (1997) momentum factor, and a liquidity factor proxied by the returns of a long-short portfolio based on stocks' Amihud measures. Returns and alphas are reported as a percentage per month. The Newey-West five-lag adjusted *t*-statistics are in parentheses. [a] 5, no coverage; 4, one analyst; for the rest, divided into 1–3.

Taming the factor zoo

- Traditionally, when proposing a new factor, the authors choose some benchmark model (FF3, FF5, FF3+UMD, etc...) to examine whether existing factors can explain the anomaly...
- ... meaning that the 300 other anomalies are not taken into account when evaluating a new factor

  $\rightarrow$ Today, we will examine how machine learning algorithms can help us to determine whether a new factor is truly new or redundant
- We will focus on whether a new factor contains a marginal contribution when explaining the cross-section of returns

A wonderful example...

- Heston and Sadka (2008) introduce a seasonality factor: a univariate sort based on past performance for the given month
- Against the FF3, the risk factor is highly significant but not if including the UMD
- $\rightarrow$ the seasonality factor is, hence, redundant when controlling for momentum...
- On the other hand, Heston and Sadka (2008) proposes several different seasonality factors while Feng et al. (2020) do not mention which one they use...

Taming the factor zoo

Objective

- We are interested in a framework to systematically examine the contribution of a risk factor(s), g_t, relative to existing risk factors using appropriate inference
- More specifically, we are interested in estimating and testing the marginal contribution of g_t conditional on existing risk factors h_t
- h_t is probably a high-dimensional set of factors in which some is probably redundant...

  $\rightarrow$ can you feel the smell of model selection (LASSO)?

Recap of SDF

- First, we need to specify our target
- From the SDF lecture, the expected return of some asset is given as

$$\mathbb{E}_t[r_{i,t+1}] - r_{f,t+1} = -(1 + r_{f,t+1})\mathrm{Cov}_t[M_{t+1}, r_{i,t+1}] \tag{1}$$

- The covariance between the stochastic discount factor and the returns explains differences in expected returns!
- The SDF is typically assumed to be linear in some factors, f_t, implying that expected returns depends on sdf loading and covariance between factors and returns!

The setting

- We will consider the following model/SDF:

$$M_t = \gamma_0^{-1} - \gamma_0^{-1}\lambda_v' v_t \equiv \gamma_0^{-1}(1 - \lambda_g' g_t - \lambda_h' h_t) \tag{2}$$

  λ's are the SDF loadings, g_t is the test factors, and h_t is the potentially confounding factors

- Note that we do not assume that h_t all are useful factors. They might have a 0 loading in the SDF
- Meaning that the expected return (not measured as excess)

$$\mathbb{E}_t[R_{t+1}] = \iota_n \gamma_0 + C_g \lambda_g + C_h \lambda_h \tag{3}$$

- Our objective is related to estimate and test λ_g controlling for h_t

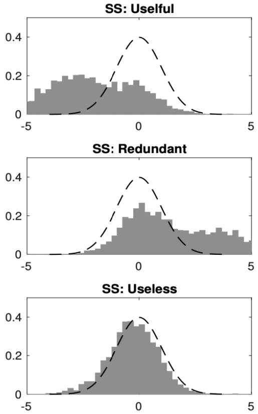
Issue with standard LASSO?

Denote by h_t a (likely high-dimensional) set of observable factor (candidate) controls, and (still) g_t the factor(s) of interest.

- The traditional LASSO is constructed with the aim of predictions
- For any finite sample, we cannot ensure that the method selects the true model
- But if important factors are excluded from the set of controls, h_t, inference about the factors of interest are erroneous

Issue with standard LASSO?

- As Feng et al. (2020) show in their appendix:



- Unless the factor is useless, using a standard LASSO for model selection, the estimates are not normally distributed

Solution to standard LASSO

- The two-pass variable selection of Belloni et al. (2014)
- This implies that a second step is added to capture missed factors from the LASSO that might introduce an OVB
- The factors not included in the two steps have minimal SDF loadings and minimal relation to the factors we want to test!

The double-selection method

- First selection: Search for factors in h_t that are useful for explaining the cross-sectional variation in expected returns. This can be done using classical
- Second selection: Search for factors in h_t that are useful for explaining the cross-sectional variation in exposures to the risk factor g_t (covariances Cov$[g_t, r_{i,t}]$). This can also be done using classical LASSO

Relationship OVB 101

- Remember that the bias of a coefficient in the present of an omitted variable is given as

$$\mathbb{E}[\hat{\beta}|X] = \beta + (X'X)^{-1}\mathbb{E}[X'Z|X]\delta. \tag{4}$$

- In the first step, we search for factors in which δ is not 0
- In the second step, we (roughly) search for factor for which $\mathbb{E}[X'Z|X]$ is non-zero

  $\rightarrow$ Using the double selection method, the two sources are minimized!

1st step

- In the first step, we run the following cross-sectional LASSO regression

$$E(R_t) = \gamma_0 + C_h \lambda_h + \varepsilon \tag{5}$$

  where C_h still is the covariance between h_t and R_t

- The Regression selects the factors being most important for explaining cross-sectional variation in expected returns
- Denote the selected factors as I_1

2nd step

- In the second step, we estimate the relationship between C_g and C_h by the following LASSO regression

$$C_{g,j} = \xi_j + C_h \chi_j' + \eta \tag{6}$$

- The second step identifies the factors whose SDF exposures are highly correlated with those of g_t
- Hence, factors that might be missed, and related to g_t, from the first step might be picked up in the second step
- Denote the union of factors from step two as I, i.e., $I_2 = \cup_{j=1}^{d} I_{2,j}$

3rd step

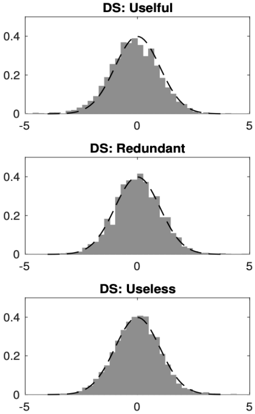
- In the third step, we run a simply cross-sectional OLS regression

$$E(R_t) = \gamma_0 + C_h \lambda_h + C_g \lambda_g + \varepsilon \tag{7}$$

- In which we restrict all factors **not** in the joint union of I_1 and I_2 to have a 0 SDF loading

Simulation results

- The simulation plot from the appendix is now given as:



- In the tables, they show that the bias is many times larger when applying a single selection relative to the double selection!

Methodology

A small note:

- Feng et al. (2020) adopt a weighting to λ_h by operator norm of the univariate betas for the factors in h_t... In other words the normalized squared betas...
- This ensures a larger penalty to variables smaller univariate betas –> helps remove spurious factors
- In practice, this can be implemented by multiplying the covariances in steps 1 and 2 by their squared mean beta across test portfolios which we normalize in the cross-section of factors

  $\rightarrow$ this implies that penalties on the coefficients are effectively smaller for high beta factors

Inference

- Feng et al. (2020) show that under some assumptions (see the appendix)

$$\sqrt{T}(\hat{\lambda}_g - \lambda_g) \xrightarrow{\mathcal{L}} N_d(0,\Pi) \tag{8}$$

  where Π is the assymptotic covariance matrix, and $\xrightarrow{\mathcal{L}}$ simply means convergence in probability (note that d is taken as the dimension of g_t)...

- The formula for the estimate of the covariance matrix is fairly long and can be seen on page 1341 in the article.
- The formula is incorporated in the Matlab script
- In the authors' replication file, they do not apply the HAC type of estimator as they write in the paper. You can easily extend the Matlab formula in our script to take this into account
- Note, that the inference procedure is valid even if we suffer from an OVB

Relationship to the three-pass estimator

- The *taming the factor zoo approach* is very much related to the three-pass estimator
- One could also use βs in Step 1 and 2 above to estimate risk premia and not SDF loadings
- This will, however, change the object such that we examine whether a new risk factor generates an adjusted return that cannot be explained by existing factors (This is written in many papers)
- $\rightarrow$ This is of course also something you can investigate!

Tuning hyperparameters

- Remember that when performing a LASSO, we have to choose how large a penalty to give coefficients. A higher λ, a more sparse model
- Gu et al. (2020) split their training sample into two: a training and a validation
- Given the in-sample focus, this does not seem like a natural way of doing things
- Instead, Feng et al. (2020) consider a K-fold cross-validation

The K-fold cross-validation

- In the K-fold cross-validation, we construct K different random subsamples
- We then perform the following steps for each subsample k:

  1. For each possible value of the hyperparameters:

     1. Train the model the K-1 other subsamples
     2. Use subsample k as "test" sample, using some evaluation metric (SSE, for instance)

- The optimal choice of the hyperparameters is the one minimizing the sum (or mean) of the evaluation metric across all K subsamples
- In this way, we mimick the out-of-sample exercise

Discussion of approach

Existing attepts...

- Feng et al. (2020) are not the first to examine the marginale contribution of introduced factors given a high dimensional
- Kozak et al. (2018) consider PCA to approximate the SDF, Kozak et al. (2020) ENet, Gu et al. (2021) autoencoder (corresponds to nonlinear PCA)
- All correspond to the first step...

  $\rightarrow$ The first step is motivated by the "oracle property" of LASSO which suggests that LASSO will recover the true model when $T \rightarrow \infty$
- As shown above this is probably not true in finite samples...
- The double-selection method can, however, can be used with the other methods such that a PCA is, for instance, used instead of LASSO

SDF loadings vs. risk premia

- The object of our exercise is to examine whether a factor can explain the cross-sectional variation
- Most of our course has examined whether a given factor is priced in the cross-section
- The two measures have, however, different economic interpretations
- The SDF loading is related to the marginal utility of investors, whereas the risk premia are related to the compensation to investors for taking some risk
- Only the SDF loading can tell us whether a factor is useful in pricing the cross-section of returns
- A nonzero risk premium can, for instance, come from correlation with the true underlying risk factors

A survey

- See Kelly and Xiu (2021) for a survey on ML in finance

Drawback

- The main drawback of the procedure is that it requires data on *all* relevant factors as opposed to just a large set of portfolios as in Giglio and Xiu (2021), which is much easier to obtain
- We still have to specify the functional form of the SDF... See Chen et al. (2020) for how to apply deep learning

Empirical findings of Giglio et. al

Data

- They consider 150 different factors from the literature constructed using the Kenneth French approach
- They consider 750 portfolios as test assets
- They consider the most recently proposed 15 factors as test factors

$\rightarrow$ They investigate whether the 15 proposed factor has a marginal contribution to explaining cross-sectional variation in the 750 portfolios conditional on the 135 other factors

The first step

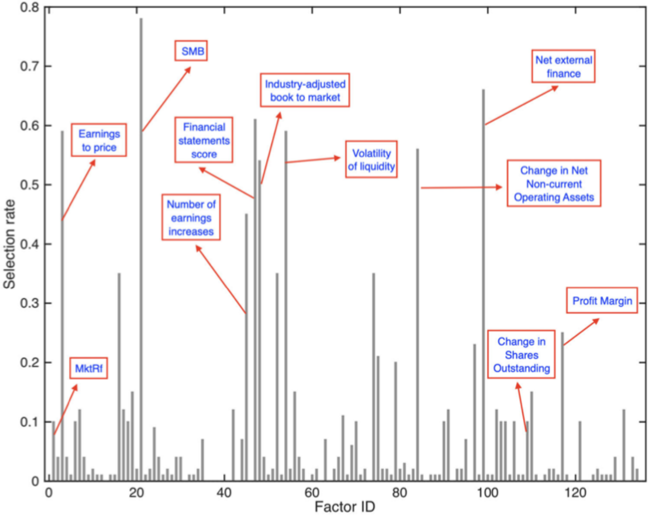
- Feng et al. (2020) identifies four factors in the linear SDF: SMB, net external finance, change in shares outstanding, and profit margin
- The question is naturally how robust the results are towards sample selection in the CV step?

Robustness first step

- To tune the hyperparameters, Feng et al. (2020) consider a cross-validation method which randomly generates K (in their case 10) subsamples
- The choice of the optimal hyperparameters, hence, depends on how the computer allocate observations into the subsamples, i.e., the random number generator
- To examine the robustness towards the random number generator, Feng et al. (2020) consider 200 different seedings
- If LASSO is perfect at selecting the true model, the output would be the same for all seedings

Robustness first step

- The following plot shows the inclusion frequency for the various factors in the first step



Robustness first step

- The LASSO is not perfect!
- Most factors is selected between 1% to 20% of the different seedings

Testing for newly introduced factors

- The table below shows their results:

		(1) DS		(2) SS		(3) FF3	
id	Factor Description	λ_s (bp)	tstat (DS)	λ_s (bp)	tstat (SS)	λ_s (bp)	tstat (OLS)
136	Cash holdings	−34	−0.42	15	0.17	10	0.54
137	HML Devil	54	1.04	−13	−0.25	−100	−2.46**
138	Gross profitability	20	0.48	3	0.06	23	2.00**
139	Organizational Capital	28	0.92	−1	−0.03	20	1.91*
140	Betting Against Beta	35	1.45	38	1.50	36	2.25**
141	Quality Minus Junk	73	2.03**	4	0.11	39	3.10***
142	Employee growth	43	1.36	−4	−0.12	−12	−0.89
143	Growth in advertising	−12	−1.18	0	0.03	12	1.32
144	Book Asset Liquidity	40	1.07	5	0.12	20	1.59
145	RMW	160	4.45***	15	0.41	20	1.80*
146	CMA	38	1.10	0	0.01	3	0.28
147	HXZ IA	51	2.11**	5	0.21	21	1.94*
148	HXZ ROE	77	3.37***	23	0.83	33	2.92***
149	Intermediary Risk Factor	112	2.21**	60	1.19	4	0.08
150	Convertible debt	−15	−1.36	−39	−3.22***	26	3.32***

$\rightarrow$ Only a handful of the factors are significant when controlling for existing factors

- Applying the FF3 model as benchmark delivers opposite conclusions for a handful of the factors

Conclusion

- The findings are robust towards changes in the tuning parameters even though the benchmark model changes substantially
- The authors suggest examining the marginal contribution when proposing a new factor
- They see this as a way to bring discipline when proposing new factors

Empirical illustration

Object

- Let us try to use the methodology to examine whether macroeconomic uncertainty exposure is priced in the cross-section of US stocks when controlling for existing factors
- Meaning whether Macroeconomic uncertainty exposure has a marginal contribution to explain cross-sectional variation relative to existing factors

Data

- We will consider the sample period from 1991 to 2019 as examined in the previous mlx file
- The test assets consist of the 202 portfolios examined in the three-pass estimator
- The confounding factor will consist of the 153 factors from the global factor data constructed by Jensen et al. (2021) in addition to the market factor from Kenneth French (I am unsure on whether the market is included in Jensen et al. (2021)). The factors are constructed as capped value-weighted
- To ease the computational burden, we will consider a 5-fold CV and only include a single seed. The code is, however, written such that you can easily change these choices

Factor construction

- Remember, from **?** that the MU exposure was estimated using the following equation using 60 months of rolling window

$$
\begin{aligned}
R^{e}_{i,t} = {} & \alpha_i + \beta_{MU,i} MU_t + \beta_{MKT,i} MKT_t + \beta_{SMB,i} SMB_t \\
& + \beta_{HML,i} HML_t + \beta_{UMD,i} UMD_t + \beta_{LIQ,i} LIQ_t \\
& + \beta_{IA,i} R_{IA,t} + \beta_{ROE,i} R_{ROE,t} + \varepsilon_{i,t}
\end{aligned} \tag{9}
$$

Factor construction

- We then construct the factor following the ones from the Kenneth French data library:
- We independently sort stocks based on size (ME) and Macroeconomic uncertainty exposure (MU), where we consider two size groups determined using the median NYSE size as the breakpoint and three momentum groups using breakpoints determined from the 30th and 70th NYSE percentiles
- The intersections provide us with six portfolios: Big Low exposure (BL), Big Neutrals (BN), Big High exposure (BH), Small Low exposure (SL), Small Neutrals (SN), and Small high exposure (SH)
- The macroeconomic uncertainty factor is then constructed from the six portfolios as follows

$$MU = \frac{1}{2}\left[SH + BH\right] - \frac{1}{2}\left[SL + BL\right], \tag{10}$$

Standard approaches

- The table below presents the alphas for different standard asset pricing benchmark models:

	FF3	FF3+UMD	FF5
α	-0.25 [-2.19]	-0.31 [-2.40]	-0.14 [-1.25]

$\rightarrow$ whether or not the anomaly can be explained by existing risk factors crucially depends on the benchmark model!

First step

- Of the 154 factors, the first step LASSO estimator selects 6
- Implying that the LASSO estimator choose a relatively sparse model
- This results depend on the choice of factors, test assets, the K in K-fold CV, and the seeding

Second step

- The second step LASSO estimator selects three different factors: market, ivol from ff3 residuals, and the ami 126 day
- The number for the second step is substantially lower than in Feng et al. (2020)
- This might be due to either:
  1. The cross-section of factors do not span the entire factor cross-section
  2. The MU factor is substially different from the rest of the factors
  3. The covariance between MU factor and the test assets is more or less completely spanned by existing factors
- To examine the last point, we can run the post-LASSO regression for the second step

	Intercept	ami	ivol	mkt	R^2
T-stat	[5.23]	[-9.84]	[-15.13]	[-9.20]	76.01%

Third step

- The test delivers a t-statistic of roughly 0 suggesting that the marginal contribution of the MU factor is nearly non-existing
- The SDF loadings are of the MU factor is, hence, spanned by existing factors
- You should, however, note that this result is probably highly dependent on many choices that I made for you

About projects

The website of Jensen et. al

- The empirical illustration from before applied the factors from Jensen et al. (2021)

- The website https://jkpfactors.com contains 153 factors for a widely selection of countries

  $\rightarrow$ The factor can be applied in many different asset pricing studies!

- You can, for instance, examine

  1. The findings of Dong et al. (2022) in an international context
  2. Whether the difference in factor performance across countries can forecast exchange rates
  3. Is there a lead-lag relationship in factor performance as examined by Rapach et al. (2013) for market returns?
  4. etc...

About projects

- Estimate and test the risk premia on one or more new risk factors, robust to OVB and measurement errors, possibly comparing to existing methods.
- Revisit fundamental risk factors, testing them and estimating their risk premia using the three-pass methodology.
- Estimate and test risk premia in other asset classes, for instance within the bonds or currency sphere.
- Estimate and test risk premia in international markets using global factors and returns (see e.g. Kenneth French's website for some data or Bryan Kelly's website for awesome data).
- Perform the study of Feng et al. (2020) for another country. For instance, Kenneth French has 150+ portfolios for European and Japanese stocks

About projects

- Adjust the analysis of Feng et al. (2020) for the publication lag. For instance, examining whether the conclusion from the 14 newly discovered factors changes if applied to the in-sample period of the 14 studies. Feng et al. (2020) has published their data on the Journal of Finance site for the article if you want to examine this
- Test the strength of existing or new risk factors via $\hat{W}$ and R_g^2. Maybe do it over different subsamples or on a rolling bias?
- $\Rightarrow$ Last, you can simply be creative

References

Bali, T. G., S. J. Brown, and Y. Tang (2017): "Is economic uncertainty priced in the cross-section of stock returns?" *Journal of Financial Economics*, 126, 471–489.

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 81, 608–650.

Chen, L., M. Pelger, and J. Zhu (2020): "Deep learning in asset pricing," *Available at SSRN 3350138*.

Dong, X., Y. Li, D. E. Rapach, and G. Zhou (2022): "Anomalies and the expected market return," *The Journal of Finance*, 77, 639–681.

Feng, G., S. Giglio, and D. Xiu (2020): "Taming the Factor Zoo: a test of new factors," *The Journal of Finance*, Forthcoming.

Giglio, S. and D. Xiu (2021): "Asset pricing with omitted factors," *Journal of Political Economy*, Forthcoming.

Gu, S., B. Kelly, and D. Xiu (2020): "Empirical asset pricing via Machine Learning," *Review of Financial Studies*, forthcoming.

——— (2021): "Autoencoder asset pricing models," *Journal of Econometrics*, 222, 429–450.

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

Jensen, T. I., B. Kelly, and L. H. Pedersen (2021): "Is there a replication crisis in finance?" *Journal of Finance*, Forthcoming.

Kelly, B. T. and D. Xiu (2021): "Factor Models, Machine Learning, and Asset Pricing," *Machine Learning, and Asset Pricing (October 15, 2021)*.

Kozak, S., S. Nagel, and S. Santosh (2018): "Interpreting factor models," *The Journal of Finance*, 73, 1183–1223.

——— (2020): "Shrinking the cross-section," *Journal of Financial Economics*, 135, 271–292.

Rapach, D. E., J. K. Strauss, and G. Zhou (2013): "International stock return predictability: What is the role of the United States?" *The Journal of Finance*, 68, 1633–1662.