Economics of Predictability in Machine Learning

Svetlana Bryzgalova

Oxford Machine Learning Summer School

ML x Finance

August 2022

Time Series Predictability

A bit of history

What is predictability and why it matters?

A stock return r_{t+1} is predictable by a factor f_t if

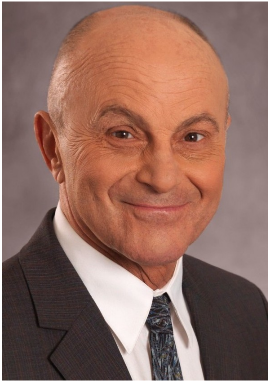
$$\mathbb{E}_t[r_{t+1}|I_t \backslash f_t] \neq \mathbb{E}_t[r_{t+1}|I_t]$$

.

Predictability of stock market returns is important for both researchers and practitioners:

- optimal asset allocation, portfolio choice
- welfare implications
- understanding the source of risks in the economy
- developing suitable structural models

Efficient Market Hypothesis



Eugene Fama
Nobel Prize in Economics, 2013

Weak form of market efficiency states that market prices already reflect all past publicly available information, hence it should not **systematically** bring additional **riskless/risk-adjusted** profits.

Up until 80s any sort of asset prices predictability (both fundamental and technical models) above transaction costs level was considered a violation of the EMH.

Basic test for market efficiency:

$$r_{t+1} = \alpha_0 + \rho f_t + \epsilon_{t+1}$$

$H_0 : \ \rho = 0$


CHALLENGE ACCEPTED

A simple test for predictability

Regression of returns on lagged returns (annual data, 1927-2012): $r_{t+1} = a + br_t + \epsilon_{t+1}$

Asset class	b	$t(b)$	R^2
Stock	0.04	0.33	0.002
T-bill	0.91	19.5	0.83
Excess stock return	0.04	0.39	0.00

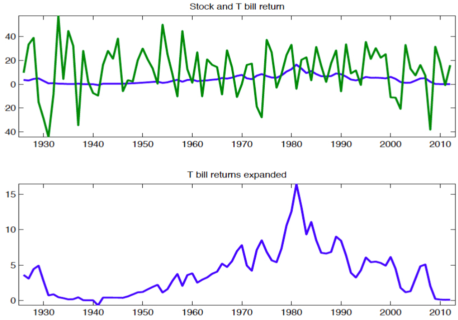


Figure 1: Historical returns on stocks and T-bills.

Market predictability

Let's use price-dividend ratio as a predictor:

Horizon k	b	$t(b)$	R^2
1 year	3.8	2.6	0.09
5 years	20.6	3.4	0.28

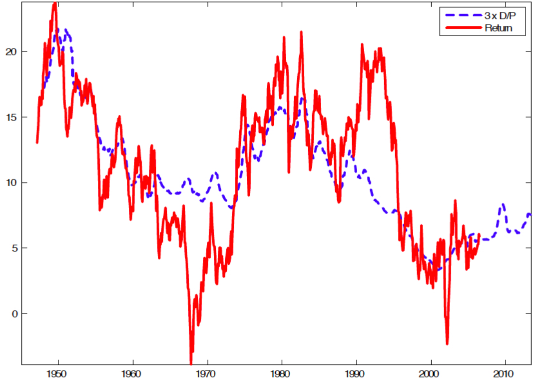


Figure 2: 7year market returns and price-dividend ratio. Source.

Understanding the null hypothesis

Early evidence for predictability of the market returns was interpreted as the violation of EMH.

$$r_{t+1} = \alpha_0 + \rho f_t + \epsilon_{t+1}$$

Inference on ρ, however, is done under a joint hypothesis:

1. market efficiency
2. linearity of the predictive model
3. constant expected returns
4. additional technical assumptions

It is now wildly accepted that **expected returns are time-varying**:

- $\mathbb{E}_t[r_{t+1}] = \alpha_0 + \rho f_t \neq \alpha_0$
- time-varying risk aversion
- long-run consumption risk
- time-varying opportunities for risk sharing (e.g. the impact of the housing collateral)
- market predictability is not the evidence for arbitrage!

A simple model for stock returns

Let's start with the definition for gross stock returns:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}$$

Log-linearizing this identity:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{\left(1 + \frac{P_{t+1}}{D_{t+1}}\right)\frac{D_{t+1}}{D_t}}{\frac{P_t}{D_t}}$$

$$r_{t+1} = \log\left(1 + e^{pd_{t+1}}\right) + \Delta d_{t+1} - pd_t$$

$$r_{t+1} \approx \log\left(1 + e^{pd}\right) + \frac{e^{pd}}{1 + e^{pd}}\left(pd_{t+1} - pd\right) + \Delta d_{t+1} - pd_t$$

$$r_{t+1} \approx \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} - (p_t - d_t)$$

where $\rho = \frac{1}{1+\frac{D}{P}} \approx 0.96$ (for annual data, $P/D \approx 20$).

Present value relationship

Start with the return identity:

$$r_{t+1} \approx \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} - (p_t - d_t)$$

Solve forward to express the present value relationship:

$$pd_t \approx \rho \times pd_{t+1} + \Delta d_{t+1} - r_{t+1}$$

$$pd_t \approx \sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j} - \sum_{j=1}^{k} \rho^{j-1} r_{t+j} + \rho^k \left(pd_{t+k}\right)$$

when $\rho^k \left(pd_{t+k}\right) \to 0$ (transversality condition, no rational bubbles)

$$pd_t \approx \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$$

Ex ante,

$$pd_t \approx \mathbb{E}_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - \mathbb{E}_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$$

Conditional on the fact that expected returns and dividend growth are both stationary, if there is a deviation in the price-dividend ratio from its long-run mean, **it should forecast**

- **future returns** $\Longrightarrow$ discount rate shocks,
- **future dividend growth**: $\Longrightarrow$ cash flow news,
- or both

Big data - big choice

Last decades saw a tremendous increase in data availability, computational power and statistics development:

- Many markets, many asset classes
- Powerful software allows to easily run thousands of regressions in minutes
- High frequency data
- Technical analysis
- New datasets
- Linear/nonlinear models, various inference procedures lead to the question of model selection and its validation
- the battery of machine learning techniques

The search for predictability continues...

Price-dividend ratio was not the only statistically and economically significant predictor:

- earnings-price ratio: Lamont (1998)
- consumption-wealth ratio: Lettau and Ludvigson (2001)
- labour income-to-consumption: Santos and Veronesi (2006)
- cross-sectional price of risk: Polk, Thompson, and Vuolteenaho (2006)
- housing collateral ratio: Lustig and S. Van Nieuwerburgh (2005)
- short term interest rates: Fama and Schwert (1977)
- credit spreads: Keim and Stambaugh (1986)
- the term structure slope: Campbell (1987)
- stock volatility: French, Schwert and Stambaugh (1987)
- equity share of new issuance: Baker and Wurgler (2000)
- aggregate short interest: Lamont and Stein (2004)
- investor sentiment: Baker and Wurgler (2006)

Furthermore, long-term returns are much more predictable:

$$\sum_{j=1}^{H} r_{t+1} - r = \kappa_r^H (dp_t - \bar{dp}) + \epsilon_{t,t+H}^r$$

with $\kappa_r^H > \kappa_r$.

Prediction zoo

Abundant empirical predictability not only for the market, but also for the performance of particular popular trading strategies/portfolios.

Novy-Marx (2014) identifies a series of new important predictors for the market and a set of popular trading strategies:

- the party of the sitting president
- the weather
- global warming
- the El Ni-ño phenomenon
- sunspot activity
- the conjunctions of the planets

Not only these results are statistically significant, quite often there is a plausible story, explaining the mechanism!

Stock market and the weather

Kamstra, Kramer, and Levi (2003): higher returns in colder months

- Higher levels of depression in Autumn (Seasonal Affective Disorder, SAD)
- Depression could lower risk appetite of the investors, and lead to lower prices
- Therefore, yields in winter become higher
- In spring, the mechanism is reversed

Cai and Wei (2006): "lower temperature can lead to aggression... [which] could result in more risk-taking... We therefore expect lower temperature to be related to higher stock returns."

But...

Hirshleifer and Shumway (2003): "psychological evidence and casual intuition predict that sunny weather is associated with upbeat mood", "sunshine is strongly significantly [positively] correlated with stock returns."

Stock and the weather

Novy-Marx (2014):
Cold weather in Manhatten predicts not just market returns, but also many trading strategies: market, but also for small cap strategies, value strategies, and strategies based on long run reversals, asset growth, and asset turnover .

Hot weather in Manhatten leads to abnormally high performance of many earnings related anomalies, including those based on return-on-assets, earnings-to-price, gross margins, financial strength, etc.

Problem:

not only weather from NYC predicts traders's mood and behaviour, but also that in Bozeman, Montana, or Hawaii.

Celestial powers

Yu, Zheng, and Zhu (2006): "...since psychological studies associate **full moon phases** with depressed mood, this study hypothesizes that stocks are valued less and thus returns are lower during full moon periods."

Contradicts Kamstra, Kramer and Levis (2003) who show that investor depression leads to an *increase* in expected returns.

Novy-Marx (2004): focus on celestial angles and sunspots

- "The aspects of Mercury and Venus with the outer planets appear particularly important for the performance of anomalies, predicting the returns of the market, and strategies based on market cap, book-to-market, momentum, gross profitability..."
- "... Mars disproportionately influences our animal instincts, especially aggression, which is strongly associated with risk-taking."
- "High levels of solar activity seem to inhibit investors capacity to process information, reducing the rate at which news gets incorporated into prices. This increases the profitability of strategies that exploit slow adjustments of prices to fundamentals."
- "... both Black Monday (October, 1987) and the start of the great recession (2007) came at minimums in the solar cycle, times of negligible sunspot activity"

What's next in prediction?

Models and inference

Lars P. Hansen
Nobel Prize in Economics, 2013

"I view the work I've done related to statistics and economics as roughly speaking, how to do something without having to do everything. So economic models – how any model by definition isn't right.

When someone just says, 'Oh, your model is wrong.' That's not much of an insight. What you want to know is, is wrong in important ways or wrong in ways that are less relevant? And you want to know what does the data really say about the model?"

Concerns about predictors

There are many concerns arising around predictive models:

- In-sample vs out-of-sample performance
- Persistency and cyclicality of regressors
- Small sample (not just the number of time series observations!)
- Instability in the linear relationship, structural breaks
- Model selection
- P-hacking, multiple testing

Two types of solutions:

- imposing a structural model on the reduced-form specification
- proper statistical tools
- **imposing economically motivated structure/constraints**

Cross-Sectional Predictability

Fundamental theorem of asset pricing

Harrison and Kreps (1979), Hansen and Richard (1987):

- In complete markets under no arbitrage there exists a unique SDF that prices all the assets in the economy.
- Under imcomplete markets under no arbitrage, there exist multiple SDF that price all the assets in the economy.

Note, the general result is applied to multi-period economies, continuum of states, etc...

Asset returns are determined by their exposure to the pricing kernel and the price of risk:

$$\mathbb{E}\left[M_{t,t+1} r^e_{t+1}\right] = 0$$

$$\mathbb{E}\left[R^e_{t+1}\right] = -\frac{cov(M_{t,t+1}, R^e_{t+1})}{\mathbb{E}\left[M_{t,t+1}\right]} = \frac{cov(M_{t,t+1}, R^e_{t+1})}{var(M_{t,t+1})} \times \left(-\frac{var(M_{t,t+1})}{\mathbb{E}\left[M_{t,t+1}\right]}\right) = \beta \times \lambda_M$$

Any asset pricing model is tested on whether it can explain the **cross-section of asset returns**

Typical way of estimating: GMM or Fama-MacBeth regressions.

50 years of empirical asset pricing in a nutshell

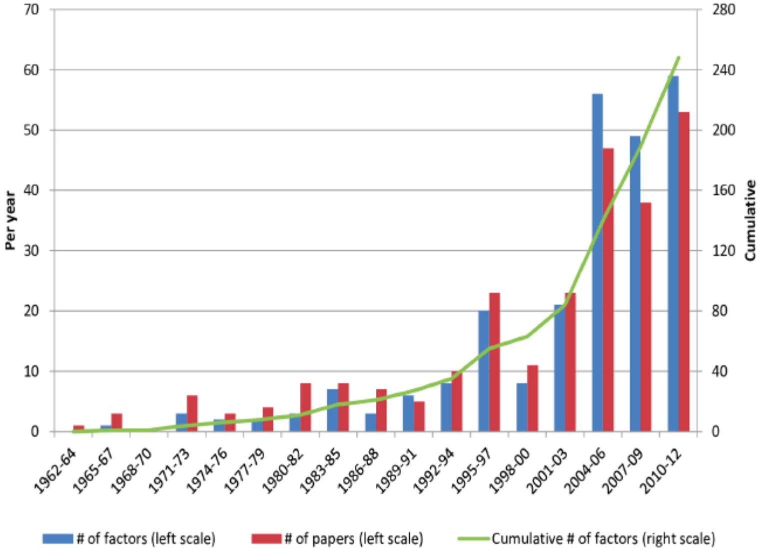
- The real implication of any asset pricing model is not how much of the returns time series in can explain, but how well it handles the cross-section of asset returns.
- Differences in exposure to systematic risk should justify differences in risk premia across various assets
- Throughout the years there has been accumulated evidence for a variety of factors being "priced".
- Why so many?

What prices the cross-section of stock returns?

Factor classification for the cross-section of stock returns, Harvey, Liu and Zhu (2016)

Risk classification		Description	Examples
Common (113)	**Financial** (46)	Proxy for aggregate financial market movement, including market portfolio returns, volatility, squared market returns, etc.	Sharpe (1964): market returns; Kraus and Litzenberger (1976): squared market returns
	Macro (40)	Proxy for movement in macroeconomic fundamentals, including consumption, investment, inflation, etc.	Breeden (1979): consumption growth; Cochrane (1991): investment returns
	Microstructure (11)	Proxy for aggregate movements in market microstructure or financial market frictions, including liquidity, transaction costs, etc.	Pastor and Stambaugh (2003): market liquidity; Lo and Wang (2006): market trading volume
	Behavioral (3)	Proxy for aggregate movements in investor behavior, sentiment or behavior-driven systematic mispricing	Baker and Wurgler (2006): investor sentiment; Hirshleifer and Jiang (2010): market mispricing
	Accounting (8)	Proxy for aggregate movement in firm-level accounting variables, including payout yield, cash flow, etc.	Fama and French (1992): size and book-to-market; Da and Warachka (2009): cash flow
	Other (5)	Proxy for aggregate movements that do not fall into the above categories, including momentum, investors beliefs, etc.	Carhart (1997): return momentum; Ozoguz (2008): investors beliefs
Individual (202)	**Financial (61)**	Proxy for firm-level idiosyncratic financial risks, including volatility, extreme returns, etc.	Ang, Hodrick, Xing and Zhang (2006): idiosyncratic volatility; Bali, Cakici and Whitelaw (2011): extreme stock returns
	Microstructure (28)	Proxy for firm-level financial market frictions, including short sale restrictions, transaction costs, etc.	Jarrow (1980): short sale restrictions; Mayshar (1981): transaction costs
	Behavioral (3)	Proxy for firm-level behavioral biases, including analyst dispersion, media coverage, etc.	Diether, Malloy and Scherbina (2002): analyst dispersion; Fang and Peress (2009): media coverage
	Accounting (86)	Proxy for firm-level accounting variables, including PE ratio, debt to equity ratio, etc.	Basu (1977): PE ratio; Bhandari (1988): debt to equity ratio
	Other (24)	Proxy for firm-level variables that do not fall into the above categories, including political campaign contributions, ranking-related firm intangibles, etc.	Cooper, Gulen and Ovtchinnikov (2010): political campaign contributions; Edmans (2011): intangibles

Factor production mill



Famous cross-sectional predictors: firm characteristics

Cross-sectional predictability: return is higher/lower for one stock relative to the other.

- relative difference, not actual positive/negative rates!
- long-short strategies

Value vs Growth:

- stocks with low book-to-market ratio (growth companies) tend to have lower returns compared with those with low B/M ratio (value companies) $\Rightarrow$ **Value premium**
- HML strategy: sort stocks based on their BM ratio, form a portfolio by buying **high** B/M stocks and selling **low** B/M stocks.

Big vs Small:

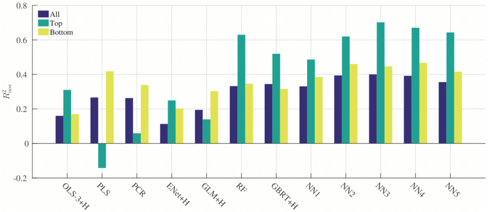
- stock with small market capitalisation (small caps) tend to have higher returns than the companies with large market capitalisation (large caps) $\Rightarrow$ **Size premium**
- SMB strategy: sort stocks based on their size, form a portfolio by buying **small** cap stocks and selling **big** cap stocks.

Gu, Kelly, Xiu (2020): Empirical Asset Pricing via ML

A wide range of standard ML algorithms, key predictors: firm characteristics

Table 1
Monthly out-of-sample stock-level prediction performance (percentage R^2_{oos})

	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 1,000	−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 1,000	−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



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 1 to 5 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 only the top-1,000 stocks or bottom-1,000 stocks by market value. The lower panel provides a visual comparison of the R^2_{oos} statistics in the table (omitting OLS because of its large negative values).

Note that a high predictive R2 does not necessarily imply good investment opportunities (Sharpe ratio, alphas, etc), nor does it guarantee asset pricing restrictions on the SDF, etc.

Which variables are important? Freyberger, Neuhier, Weber (2020)

Flexible additively nonparametric form of the impact of characteristics on asset returns:

$$r_{i,t} = m_{t,1}(C_{1,i,t-1}) + m_{t,2}(C_{2,i,t-1}) + \epsilon_{i,t}$$

Approximate nonlinear functions with splines, and impose variable selection via grouped lasso (GL):

$$\hat{\beta}_1 = \arg\min_{b_{s,k}} \sum_{i=1}^{N} \left(r_{i,t} - \sum_{s=1}^{S} \sum_{k=1}^{L+2} b_{s,k} p_k(C_{s,i,t-1}) \right)^2 + \lambda_1 \sum_{s=1}^{S} \left(\sum_{k=1}^{L+2} b_{s,k}^2 \right)^{1/2}$$

Define characteristic weights, inversely proportional to the size of chosen parameters and run adaptive GL again:

$$\hat{\beta}_2 = \arg\min_{b_{s,k}} \sum_{i=1}^{N} \left(r_{i,t} - \sum_{s=1}^{S} \sum_{k=1}^{L+2} b_{s,k} p_k(C_{s,i,t-1}) \right)^2 + \lambda_2 \sum_{s=1}^{S} \left(w_s \sum_{k=1}^{L+2} b_{s,k}^2 \right)^{1/2}$$

where $w_s = \left(\sum_{k=1}^{L+2} \hat{\beta}_{1,s,k}^2 \right)^{-1/2}$

Again, this does not lead to great investment decisions, nor does it impose a no-arbitrage constraint.

Kelly, Su and Pruitt (2019): IPCA

Characteristics of the assets are **informative** of their factor loadings on the common sources risk:

$$\underbrace{r_t^{ex}}_{N\times 1} = \underbrace{\beta_{t-1}}_{N\times K}\underbrace{F_t}_{K\times 1} + \epsilon_t$$

where F_t are latent common factors, and conditional factor loadings β_{t-1} are functions of characteristics C_{t-1}:

$$\underbrace{\beta_{t-1}}_{N\times K} = \underbrace{C_{t-1}}_{N\times L}\underbrace{\Gamma}_{L\times K}$$

The main contribution of the paper lies in putting forward **a coherent framework** that is flexible and easy to estimate, but retains the economic intuition.

IPCA procedure

- Definition of the estimator:

$$\min_{\Gamma,F} \frac{1}{NT}\sum_t \left(r_t^{ex} - \beta_{t-1}F_t\right)' \left(r_t^{ex} - \beta_{t-1}F_t\right) \quad \text{s.t. } \beta_{t-1} = C_{t-1}\Gamma$$

- F.O.C. w.r.t F_t:

$$F_t = \left(\beta_{t-1}'\beta_{t-1}\right)^{-1}\beta_{t-1}' r_t^{ex} = \left(\Gamma'\Gamma\right)^{-1}\beta_{t-1}' r_t^{ex}$$

since $\frac{C_t' C_t}{N} = I$

- Concentrating w.r.t the solution to this F.O.C.:

$$\max_{\Gamma} \; tr\left\{ \left(\Gamma'\Gamma\right)^{-1}\Gamma' \left(\frac{1}{T}\sum_t \frac{C_{t-1}' r_t^{ex} r_t^{ex\prime} C_{t-1}}{N^2} \right) \right\}$$

- Γ is given by first K eigenvectors of $Z = \frac{1}{T}\sum_t \frac{C_{t-1}'(r_t^{ex})(r_t^{ex})' C_{t-1}}{N^2}$

Simple stepwise procedure:

1. Take a set of characteristics and returns, construct matrix Z, extract K eigenv. $\hat{\Gamma}$
2. Compute $\hat{\beta}_{t-1} = C_{t-1}\hat{\Gamma}$
3. Get F_t from the cross-sectional regression of returns on betas:

$$\hat{F}_t = \left(\hat{\beta}_{t-1}'\hat{\beta}_{t-1}\right)^{-1}\hat{\beta}_{t-1}' r_t^{ex}$$

Nonlinear effects

Fan, Liao and Wang (2016): characteristics proxy stock loadings on the risk factors

$$r_{i,t}^{ex} = \sum_{k=1}^{K} \left(g_k(C_i) + \alpha_{i,k}\right) F_{t,k} + \epsilon_{i,t}$$

- betas are general nonlinear functions of characteristics
- dimensionality curse

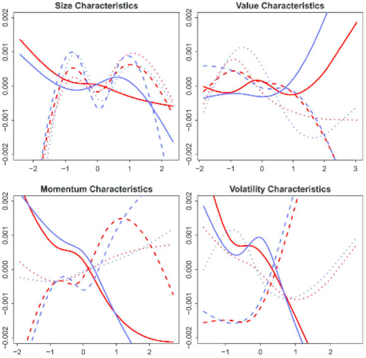


FIG. 1. *Estimated additive loading functions g_{kl}, $l = 1, \ldots, 4$ from financial returns of 337 stocks in S&P 500 index. They are taken as the true functions in the simulation studies. In each panel (fixed l), the true and estimated curves for $k = 1, 2, 3$ are plotted and compared. The solid, dashed and dotted red curves are the true curves corresponding to the first, second and third factors, respectively. The blue curves are their estimates from one simulation of the calibrated model with $T = 50$, $p = 300$.*

Structural deep learning in asset pricing

Fan, Ke, Liao, Neuhierl (2022):

$$r_{i,t} = \alpha_{i,t-1} + \beta'_{i,t-1}\lambda_{t-1} + \beta'_{i,t-1}(f_t - \mathbb{E}[f_t]) + u_{i,t}$$

where

$$\alpha_{i,t-1} = g_{\alpha,t}(x_{i,t-1}) + \gamma_{\alpha,i,t-1}, \quad \mathbb{E}[\gamma_{\alpha,i,t-1}|x_{i,t-1}, f_t] = 0$$
$$\beta_{i,t-1} = g_{\beta,t}(x_{i,t-1}) + \gamma_{\beta,i,t-1}, \quad \mathbb{E}[\gamma_{\beta,i,t-1}|x_{i,t-1}, f_t] = 0$$

Structural machine learning prediction via DNN:

$$E[r_{i,t}] = g_{\alpha_{i,t-1}}(x) + g_{risk}(x) + g_{factor}(x)$$

- period-by-period cross-sectional deep learning, followed by local PCAs
- captures time-varying features such as latent factors of the model.
- formal asymptotic theory of the structural deep-learning estimators, which apply to both in-sample fit and out-of-sample predictions.

Three Applications of ML in Finance

The plan

A brief overview of the three conceptually different projects

1. Testing model selection and sparsity in linear factor asset pricing models
   $\Longrightarrow$ "Bayesian Solutions for the Factor Zoo: We Just Ran 2 Quadrillion models"
   (with C. Julliard and J. Huang, forthcoming in the *Journal of Finance*)

2. How to build cross-sections of portfolios?
   $\Longrightarrow$ "Forest through the Trees: Building a Cross-section of Asset Returns"
   (with M. Pelger and J. Zhu, R&R at the *Journal of Finance*)

3. How to measure economic sentiment for almost 200 years?
   $\Longrightarrow$ "Economic News and the Macroeconomy"
   (with J. van Binsbergen, M. Mukhopadhyay, and V. Sharma)

Bayesian Solutions for the Factor Zoo

The big questions

Q1: Is there a single set of observable factors (a dominant model) that rationalises asset prices? I.e., does model selection make sense?

Q2: Do we really need the whole Zoo to capture priced risk?
I.e., is the "true" pricing kernel/SDF/tangency sparse or dense in observable factors?

Q3: Can we just use leading Principal Components to characterise pricing information?

Q4: Are nontradable factors (consumption, sentiment, etc.) doomed?

We consider 51 factors from the literature, yielding $2^{51} \approx 2.25$ quadrillion models,

- $\approx$ 25,000 galaxies (in stars)
- $\approx$ 15 brains (in synapses)

We examine the cornerstone assumptions of uniqueness, identification, and sparsity.

Bayesian basics I

Parameter as random variables: $\Rightarrow$ focus on their "posterior" distribution

- start with a prior about parameter value/range
- update it, given the data (posterior distribution)
- learning view of working with data

Single model estimation: we develop a Bayesian SDF estimation

✓ Estimation/Inference of price of risk and any other measure (e.g., R^2, SR, etc..)

✓ Automatically robust to weak/level factors

✓ As fast as OLS/GLS regression

✓ Cross-sectional likelihood of the model

Bayesian basics II

"All models are wrong, but some are useful."
Box (1976)

Model posterior probabilities:

$$\Pr\left(\text{model } j|\text{data}\right) = \frac{\Pr\left(\text{data}|\text{model } j\right)\Pr\left(\text{model } j\right)}{\Pr\left(\text{data}\right)}$$

where $\Pr\left(\text{data}|\text{model } 0\right) = \int \text{likelihood}(\text{data}|\text{Model } 0, \theta^{(0)})\pi(\theta^{(0)})d\theta^{(0)}$, $\theta^{(0)}$ are the model 0 parameters and π their prior distribution.

Information criteria (e.g., AIC, BIC etc.) are approximations of marginal likelihood.

Bayesian Model Averaging:

If we are interested in some Δ well-defined for every model:

$$\mathbb{E}\left[\Delta|\text{data}\right] = \sum_{j=0}^{M} \mathbb{E}\left[\Delta|\text{data}, \text{model} = j\right]\Pr\left(\text{model} = j|\text{data}\right)$$

- Sharpe ratio, risk premia, alphas, factor inclusion probability, etc

Continuous spike-and-slab & how to handle quadrillion models

Instead of estimating all the models, sample them.

1. Specify the λ prior as:
$$\lambda_j|\gamma_j, \sigma^2 \sim N(0, r(\gamma_j)\psi_j\sigma^2)$$
where $r(1) = 1$ and $r(0) = r \lll 1$, i.e. use continuous Spike-and-slab.
   - similar, integrable, posterior as before

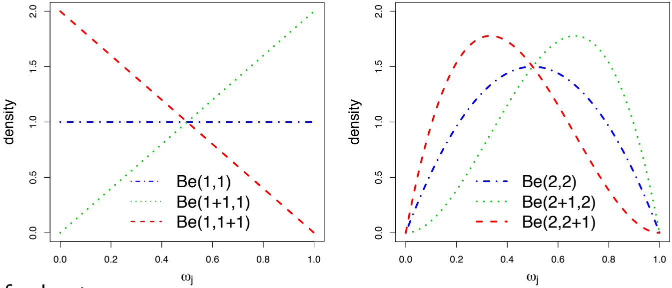
2. Treat γ_j as a parameter to be sampled,
   - posterior factor probability $= \mathbb{E}[\gamma_j|\text{data}]$
   - $\Rightarrow$ Estimate model/factor probs from posterior sampling frequency

$\Rightarrow$ Markov Chain over the space of possible models.
   - Large efficiency gains as low probability models get endogenously undersampled
   - Can construct BMA-SDF using ALL possible models

Note: BMA gives "best" predictive density (min KLIC + optimal on "average")

Sampling factors and sparsity

Factor inclusion prior describes model sparsity: $\pi(\gamma_j = 1|\omega_j) = \omega_j, \quad \omega_j \sim Beta(a_\omega, b_\omega)$.

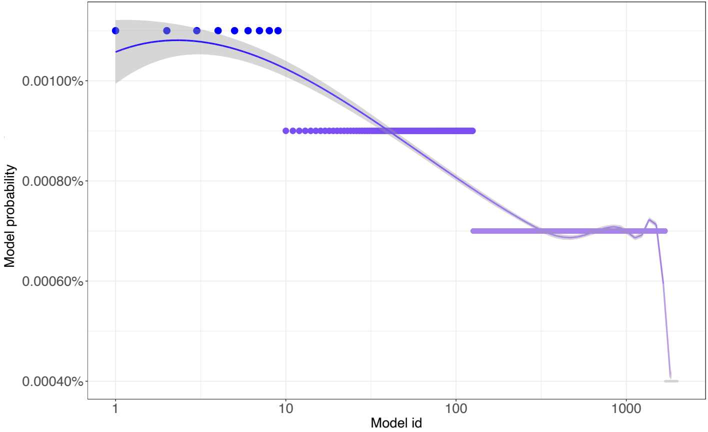


Note: beliefs about

1. Sharpe ratio achievable with one factor
2. Sharpe ratio achievable in the economy
3. Sparsity pattern of the true model

fully determine prior parameters.

Model probabilities: massive model uncertainty



Posterior model probabilities in a cross-section of 60 factors/anomalies under the total SR prior, $\sqrt{\mathbb{E}_\pi[SR_f^2|\sigma^2]} = 2$.

Note: frequentist LR test of the best performing model versus the 1000th one would yield a p-value of 30% at best

Each model prior: 4×10^{-16} (with the prior for 1000 models of 4×10^{-13})

Posterior probabilities of notable models and "robust" factors model

We have found a lot of uncertainty about the "true" model...

Posterior probabilities (%) of notable models versus robust factors

	Panel A: 3 most likely factors					*Panel B*: Most likely 5-factor model				
model: SR_{pr}:	1	1.5	2	2.5	3	1	1.5	2	2.5	3
cobalt!22 Most likely factors	17.5%	24.9%	36.0%	48.8%	59.1%	23.0%	35.3%	57.0%	77.6%	88.1%
CAPM	12.7%	12.5%	11.8%	11.3%	13.1%	11.9%	10.8%	8.0%	5.0%	3.9%
FF3	10.3%	7.9%	5.3%	3.2%	1.7%	9.6%	6.8%	3.6%	1.4%	0.5%
FF5	9.9%	7.0%	4.2%	2.1%	0.7%	9.2%	6.0%	2.8%	0.9%	0.2%
Carhart	10.2%	7.8%	5.2%	2.9%	1.3%	9.6%	6.7%	3.5%	1.3%	0.4%
q4	15.7%	17.8%	17.9%	14.9%	9.6%	14.6%	15.3%	11.9%	6.4%	2.7%
Liq-CAPM	12.5%	12.0%	10.9%	9.6%	9.0%	11.7%	10.4%	7.4%	4.3%	2.7%
FF3-QMJ	11.2%	10.1%	8.8%	7.4%	5.5%	10.4%	8.6%	5.8%	3.1%	1.5%

Posterior model probabilities for the specifications in the first column, for different prior Sharpe ratio values, computed using the Dirac spike-and-slab prior method. Panel A includes the factors BEH_PEAD, MKT, CMA*. Panel B uses the most likely 5-factor model according to the posterior probability. Factors are: MKT, MGMT, BAB, BEH_PEAD, CMA* for $SR_{pr} = 1$; STRev, BAB, BEH_PEAD, RMW*, CMA* for $SR_{pr} = 1.5$ to 2.5; BW_ISENT, BEH_PEAD, MKT*, RMW*, CMA* for $SR_{pr} = 3$. Liq-CAPM stands for the liquidity-adjusted model of Pastor and Stambaugh (2003) and FF3-QMJ corresponds for a 4-factor model of Asness, Frazzini, and Pedersen (2019). Sample: 1973:10 to 2016:12, test assets: 60 factor/anomaly portfolios.

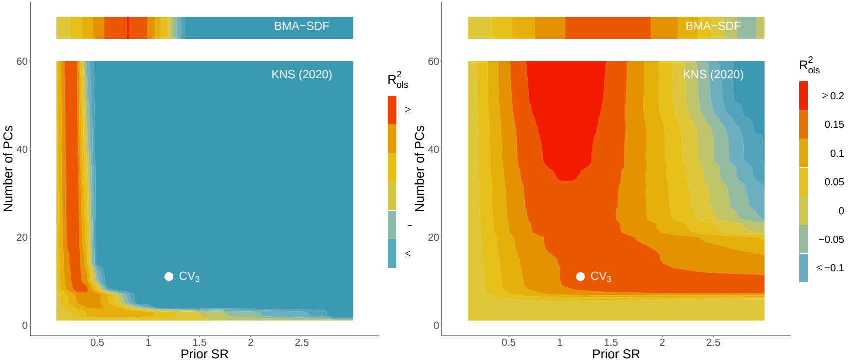
... but there is little doubt the models we've been using so far are not good enough.

Note: none of the above models is among the 2000 most likely ones.

$\Rightarrow$ need for a better model: BMA-SDF

OOS pricing: past and future

BMA-SDF vs Kozak, Nagel and Santosh (2020):



(a) Pricing 2nd half of the sample

(b) Pricing 1st half of the sample

The Big Questions & Answers

Q1: Is there a single set of observable factors (a dominant model) that rationalises asset prices? I.e., does model selection make sense?

A1: Not really. The best few thousands models (none ever proposed in the literature) are virtually the same empirically. Questioning the very idea of model selection.

Q2: Do we really need the whole Zoo to capture priced risk?
I.e., is the "true" SDF sparse or dense in observable factors?

A2: The SDF is dense in observable factors $\Rightarrow$ low-dimensional models are misspecified, and the Zoo is actually a jungle of noisy proxies of common underlying sources of risk

Q3: Can we just use leading Principal Components to characterise pricing information?

A3: Nope. There is value in observable factors above and beyond (RP-)PCs. And both are optimally combined by the BMA-SDF.

Q4: Are nontradable factors (consumption, sentiment, etc.) doomed?

A4: Not at all. Especially if we are in search of a low-dimensional model.

A3 + A4 = theory ain't dead :)

Forest through the Trees: How to build cross-sections of portfolios?

Cross-sections and Test Assets

Asset pricing has two essential elements:

1. A cross-section of test or basis assets (e.g. 25 size-value sorted portfolios)
2. A model for their pricing (e.g. the Fama-French 3-factor model)

$\Rightarrow$ Informative test/basis assets crucial for evaluating existing and building new models

Fundamental problem:

- Building cross-sections of managed portfolios conditional on stock characteristics
- Classification problem: Which stocks should be grouped together?
- Key insight: Test assets have to span the SDF (conditional on stock characteristics)

Challenges:

- Complex dependency on multiple characteristics - interaction effects
- Large number of available characteristics - curse of dimensionality
- Redundancy in characteristics: Repackaging of the same risk
- Interpretability: What features are hard to explain?

This paper: Use decision trees to build small cross-sections of informative, interpretable test assets, conditioned on many stock characteristics.

Conventional Approach: Sorting with Multiple Characteristics

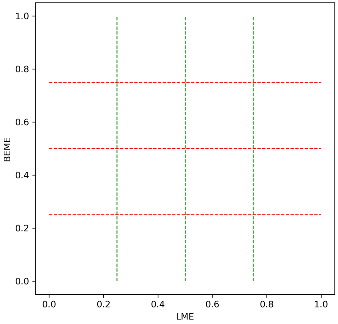


Figure 4: Unconditional quantiles of 16 double-sorted portfolios

- Sorting is routinely used to create cross-sections based on 1-2 characteristics
- More than 3 characteristics: stack several cross-sections together
  E.g., 25 size and value + 10 momentum

Conventional Approach: Sorting with Multiple Characteristics

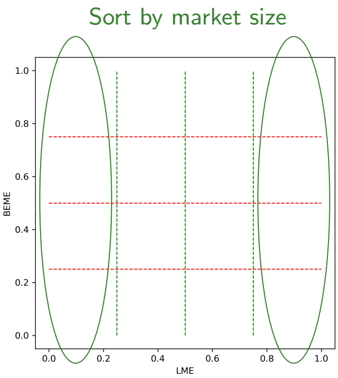


Figure 4: Unconditional quantiles of 16 double-sorted portfolios

- Sorting is routinely used to create cross-sections based on 1-2 characteristics
- More than 3 characteristics: stack several cross-sections together
  E.g., 25 size and value + 10 momentum

Conventional Approach: Sorting with Multiple Characteristics

Sort by book-to-market

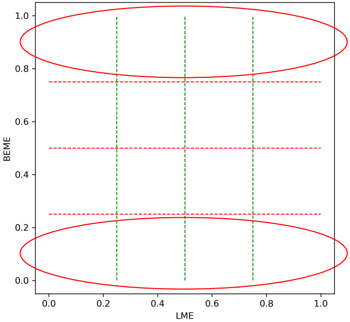


Figure 4: Unconditional quantiles of 16 double-sorted portfolios

- Sorting is routinely used to create cross-sections based on 1-2 characteristics
- More than 3 characteristics: stack several cross-sections together
  E.g., 25 size and value + 10 momentum

Conventional Approach: Sorting with Multiple Characteristics

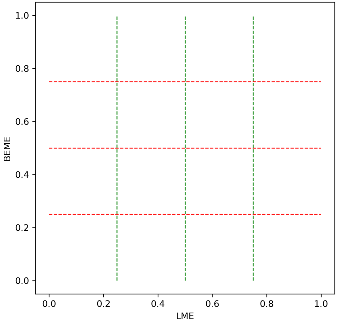


Figure 4: Unconditional quantiles of 16 double-sorted portfolios

- Sorting is routinely used to create cross-sections based on 1-2 characteristics
- More than 3 characteristics: stack several cross-sections together
  E.g., 25 size and value + 10 momentum

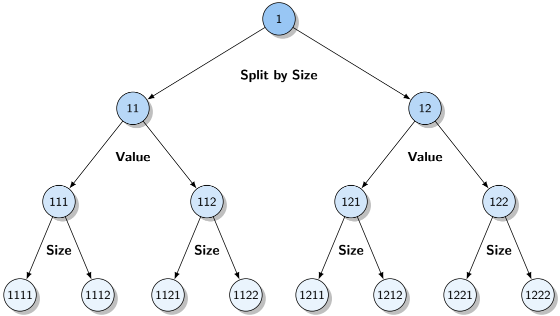
Challenges:

- Interactions: coarse or none
- Where to draw the boundaries?
- Not considering firm distribution
  $\Rightarrow$ empty and unbalanced portfolios

$\Rightarrow$ Success of simple long-short factors may be due to overly simplistic test assets.

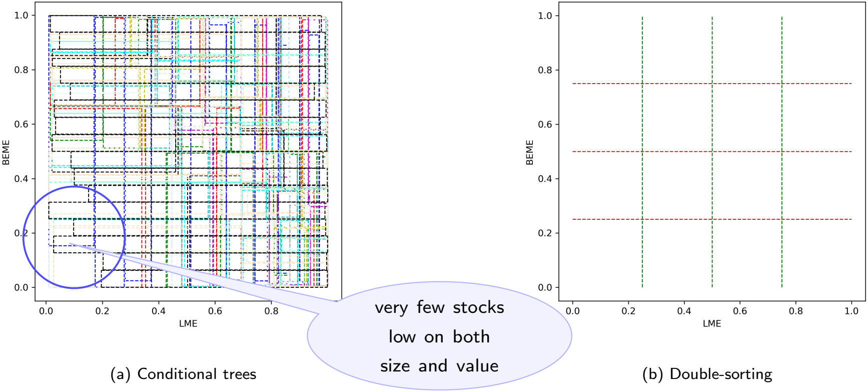
Trees and conditional splits

A tree of depth $d = 3$ is based on $k = 2$ characteristics and 50/50 splits:



- Simplest trees rely on the 50/50 split within each group of the assets
- Conditional on the order of splits, each subtree yields 2^d non-intersecting portfolios.
- AP-Trees: $k^d \times 2^d$ portfolios of all orders, with $\frac{N}{2^d}$ stocks in each independent of k.
- Tree-based portfolios are conditional splits:
  Reflect joint distribution of characteristics without the need for parametric modeling

Size and Value: Trees vs Double-Sorting



(a) Conditional trees

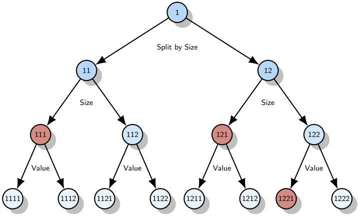
(b) Double-sorting

Figure 5: Cross-sectional quantiles of the portfolio splits based on trees and double sorting.

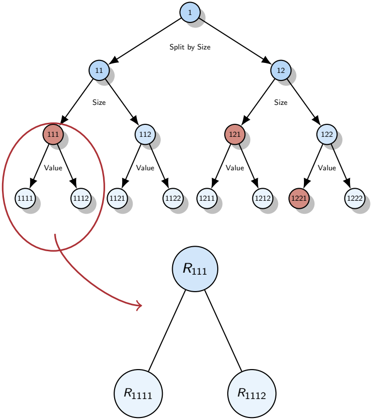
- 1296 final nodes with 3 characteristics and depth 4 (+ intermediate ones)
- Flexible set of basis functions that can easily span the SDF

⇒ Need for dimension reduction: Pruning = decision to make a split

Pruning the trees



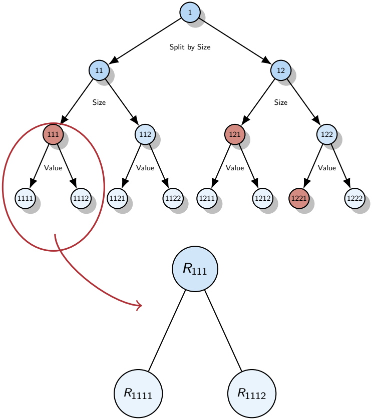
Pruning the trees



Conventional: Local bottom-up approach, e.g.

- compare returns of 1111 and 1112,
- if they are different, keep them separate.

Pruning the trees



Conventional: Local bottom-up approach, e.g.

- compare returns of 1111 and 1112,
- if they are different, keep them separate.

Problem: Selecting blocks for the SDF

- Our goal is to find the best set of basis assets
- Together they should give the highest SR
- Optimal decision depends on other nodes
- Pruning is global selection problem
- Split only if it increases the overall SR
- Bias-variance tradeoff: Higher nodes more diversified

Our approach: **Asset Pricing Pruning**

$\Rightarrow$ Recursive structure: Pruning = sparse selection on final and intermediate tree nodes

$\Rightarrow$ AP-Pruning = Selection with optimal mean–variance portfolios + shrinkage applied to all the final and intermediate nodes of tree portfolios

Asset Pricing Pruning: sparse and efficient portfolio frontier

Take all the final and intermediate nodes from all the trees:

1. Construct a portfolio frontier on training data with elastic net (under μ, Σ uncertainty):
$$\begin{aligned} \text{minimize} \quad & w^T \Sigma w + \lambda_1 ||w||_1 + \lambda_2 ||w||_2^2 \\ \text{subject to} \quad & w^T 1 = 1 \\ & w^T \mu \geq \mu_0 \end{aligned}$$
Tuning parameters: target return μ_0, sparsity lasso shrinkage λ_1, ridge shrinkage λ_2.

2. Choose robust tangency portfolio by selecting tuning parameters on the validation set.

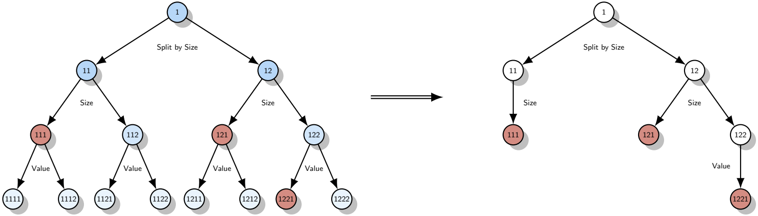


Figure 6: A single pruned standalone tree: only dark red nodes are retained.

Sharpe ratios

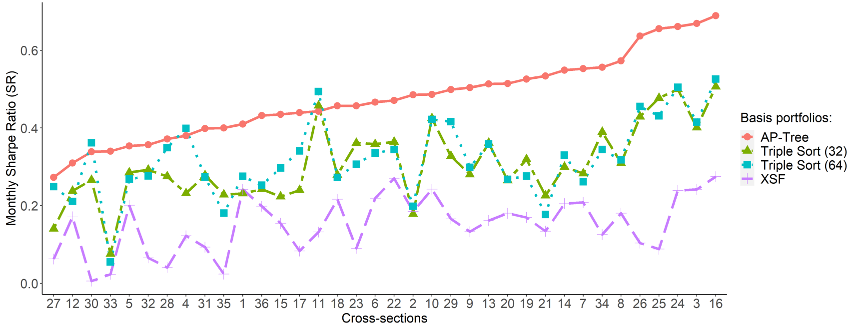


Figure 7: Monthly out-of-sample Sharpe ratios of SDFs spanned by AP-Trees and triple sorts

- Cross-sections sorted by the Sharpe ratios (SR) of AP-Trees.
  For example: cross-section 2: size, value, profitability
- Cross-section specific factors XSF= market + 3 specific factors

⇒ SDF of AP-Trees has up to 2x the SR of triple sorts and 3x SR of XSF

Usual factors cannot price AP-Trees within cross-sections

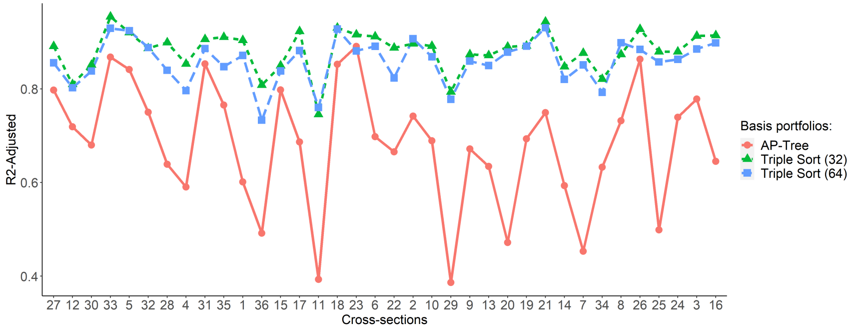


Figure 8: $XS\text{-}R^2$: Cross-sectional pricing of portfolios XS-specific factors.

- Triple-sorted portfolios are easier to price than AP-Trees for all the characteristics
- Counting metrics like $XS\text{-}R^2$ are inflated by redundant and oversimplified triple sorts

AP-Trees vs general machine learning

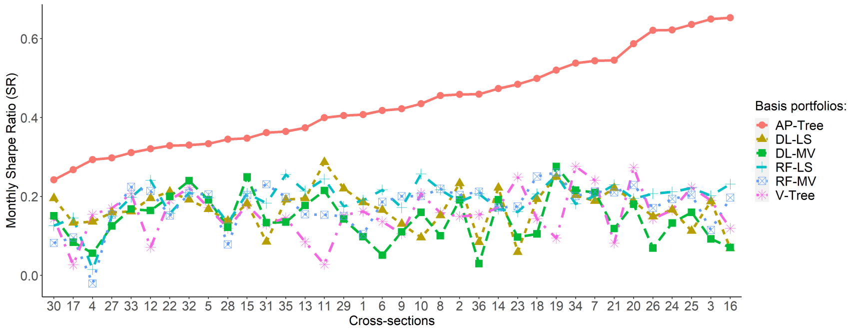


Figure 9: Monthly out-of-sample Sharpe ratios of AP-Trees vs off-the-shelf ML methods.

⇒ AP-Trees capture more pricing information than variation based (V-Trees) or forecasting based (Deep Learning, Random Forest) decile portfolios

⇒ Correct objective function: SDF spanning and not return spread or variance

Predicting returns $\neq$ building a cross-section

AP-Trees vs general machine learning

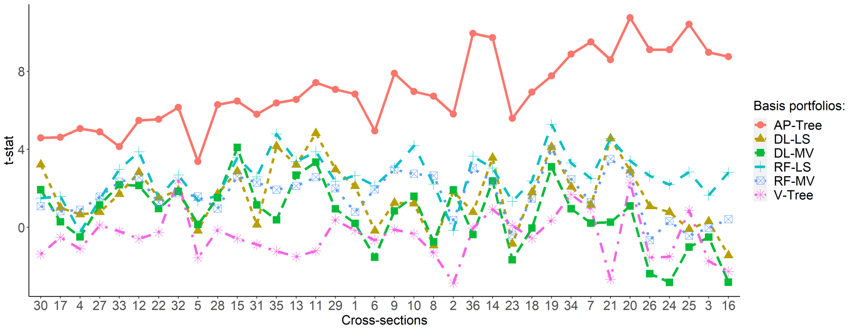


Figure 9: SDF alpha relative to Fama-French 5 factor model.

$\Rightarrow$ AP-Trees capture more pricing information than variation based (V-Trees) or forecasting based (Deep Learning, Random Forest) decile portfolios

$\Rightarrow$ Correct objective function: SDF spanning and not return spread or variance

Predicting returns $\neq$ building a cross-section

Measuring Economic Sentiment for 170 years

Motivation

- Growing empirical evidence on the role of expectations and sentiment over the business cycle: Greenwood and Hanson (2013), Lopez-Salido et al. (2017), Mian et al. (2017), Bordalo et al. (2018a,b), Bordalo et al. (2020)
- To understand the role of sentiment and the content of news, it it important to first **measure** them
- Existing surveys are very short-lived, and target very few economic objects

This paper:

- Use the biggest corpus of newspaper articles in the world to construct text-based measures of sentiment
- New (for econ and finance) state-of-art methodology to extract sentiment about a particular topic
- Local and country-level measure of sentiment, for about 170 years
- What is the information (local and global) and variation in economic news?

Dataset

- Historical collection of digitized newspapers from 1736-2020 in the USA
- 13,000 local newspapers across thousands of communities
- 193 million newspaper pages
- Around 2 billion newspaper articles (compared to 800,000 for the standard WSJ corpus)
- The corpus is **huge**! 95 times larger than all of English Language Wikipedia!
- Madison Project database for long-term gdp data

A general approach of sentiment measures

Sentiment analysis in economics/finance typically follows 3 broad steps:

- **Step 1**: Create a dictionary of relevant words for the topic
- **Step 2**: Identify the words that are positive or negative
- **Step 3**: Apply this dictionary to measure the intensity of positive and negative sentiment in a particular document.

Our approach: state-of-the-art machine learning technique, pioneered in Singla and Mukhopadhyay (2022)

Step I: Automated Dictionary

Semantic information about every word is captured by it's vector representation, based on the context/embedding in which the word/phrase is used.

- Word2vec, based on neural networks
- Applies to phrases as well as words
- Outperforms existing alternatives, Mikolov et al. (2013)

Create dictionary of most similar words

- Use distances between the vector representations of words to measure their similarities

Example:

- The word **economy** is associated with the words/phrases: *economic growth, inflation, economic, recession, economic recovery, consumer spending*, etc., in our newspaper database

Step II: Automated Measure of Word Sentiment

Isolate **positive** or **negative** words to measure economic sentiment

Start with a few seed words related to sentiment:

- Positive: expansion, boom, growth, profit, optimistic, optimism, opportunity, success, successful, profitable, prosperity, profitability, bullish
- Negative: recession, bankrupt, shrinking, unemployment, loss, bankruptcy, cutback, layoff, redundancy, pessimism, contraction, unsuccessful, failure, insolvent, insolvency, bearish

Following Hamilton et al. (2016) – the gold standard to generate sentiment scores for economics and finance – we produce a continuous measure of sentiment for each word/phrase in the dictionary:

- top 10 most positive words/phrases: *successful, success, profit, opportunity, profitable, cess, proven, expansion, interview, rewarding*
- top 10 negative words/phrases: *failure, unsuccessful, insolvent, earnest, redundancy, bankruptcy, bankrupt, insolvent debtor, contraction*

Step III: Aggregating Sentiment of a Document

- Produce a vector representation of the document, as in Arora et al. (2017)
- Focus on the pages with economy-related words
- Measure the similarity between the vector representation of each newspaper page and our economic dictionary (weighted by the word sentiment score)
- Aggregate to the city/county/state/country level
- Seasonally adjust the final measure

Key Advantages

- Almost fully automated construction of topic specific dictionaries
  - can be easily applied to almost any other context/topic
  - no need to manually create the whole dictionary
- Reflects the context in which words/phrases are used
- Automatically deals with negation, a common challenge for word counts
  - good vs not good
- Automatic generation of a continuous measure of word sentiment score
  - good vs excellent
- Captures not just key words, but their synonyms and related language

Quarterly measure of economic sentiment

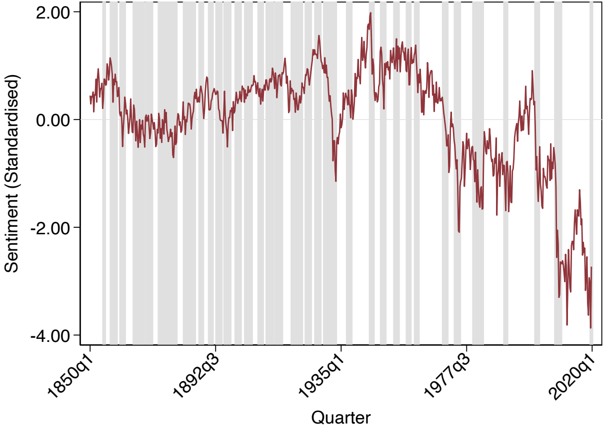


Figure 10: National Economic Sentiment: Quarterly data, 1850-2020

Monthly measure

Sentiment varies a lot, and is substantially lower during recessions.

Quarterly measure of economic and non-economic sentiment

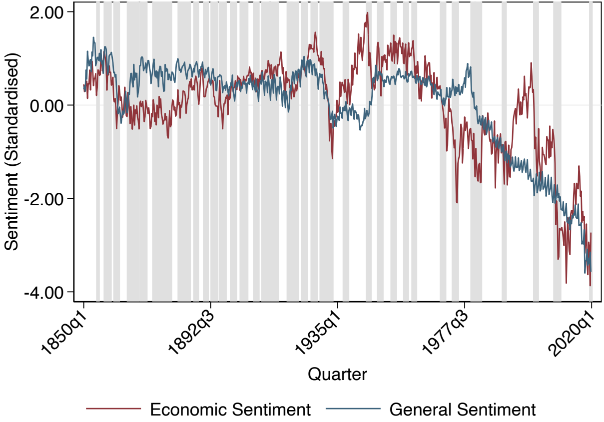


Figure 11: National Economic and Non-Economic Sentiment: Quarterly data, 1850-2020

Non-econ sentiment is different and is trending downwards from 1980.

Pure national economic sentiment

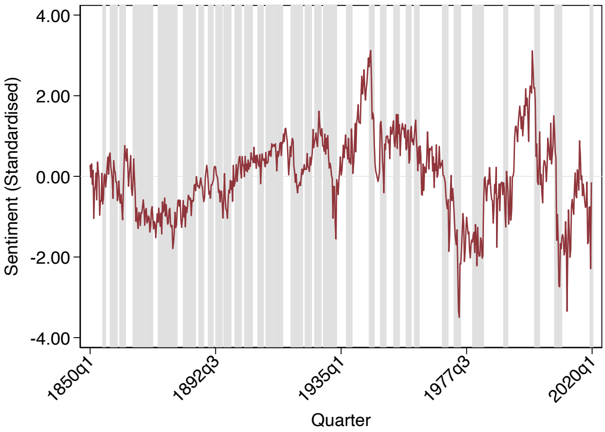


Figure 12: National Economic Sentiment (Orthogonalised): Quarterly data, 1850-2020

Orthogonalising economic w.r.t non-econ sentiment produces a stationary series
This procedure could be used to clean other existing measure of common trends

Current vs. future

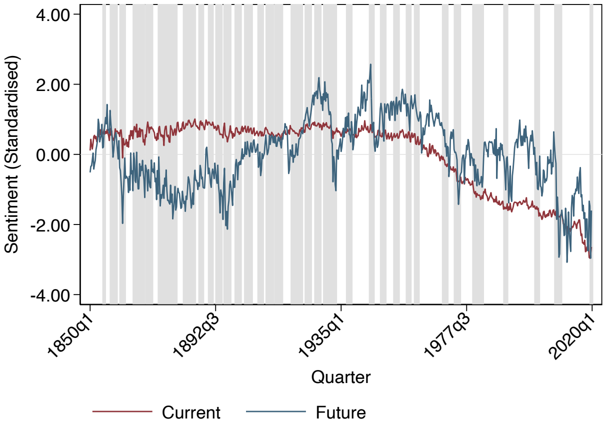
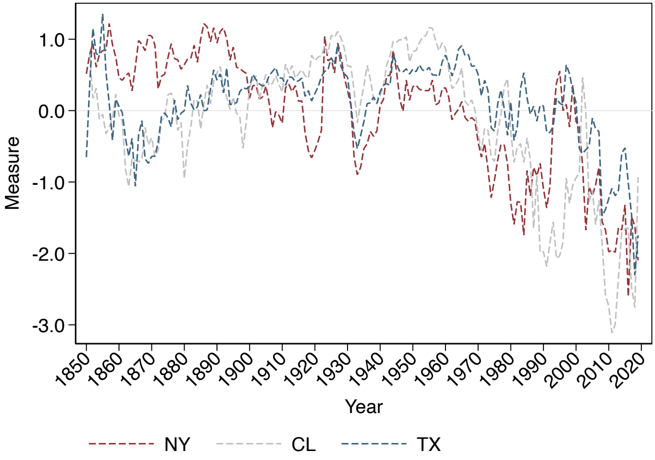


Figure 13: National Economic Sentiment Current vs. future: Quarterly data, 1850-2020

Substantial variation in the local economic sentiment

Figure 14: Local Economic Sentiment



National trend only explains around 35% of the variation in state sentiment

Measures of economic sentiment for 1850-2020 in the USA

Country level:

- sentiment has a clear business cycle pattern
- it co-moves with existing measures (Michigan survey)
- sentiment predicts GDP growth, employment, and consumption and services controlling for past fundamentals
- it leads the survey of professional forecasters

State level:

- economic sentiment predicts local GDP growth, controlling both state and country-wide fundamentals and national level of sentiment
- common trend drives only about half of the state-level sentiment
- natural measure dispersion in sentiment across states
- higher dispersion across states predicts lower GDP growth

Key methodological contribution: a state-of-the art NLP approach to measuring topic-specific sentiment, following **Singla and Mukhopadhyay (2022)**

Machine learning use in finance: new methods needed!

1. Simple prediction: find signal to predict future returns
   - linear and nonlinear regressions,
   - deep learning, lasso, random forest: **standard ML is great!**
2. Ultimate object of interest: use prediction output for
   - long-short strategies, portfolio optimization,
   - constructing SDF,
   - rebalancing, trading frictions, transaction costs, etc.

$\Longrightarrow$ Need to combine economic structure with ML.