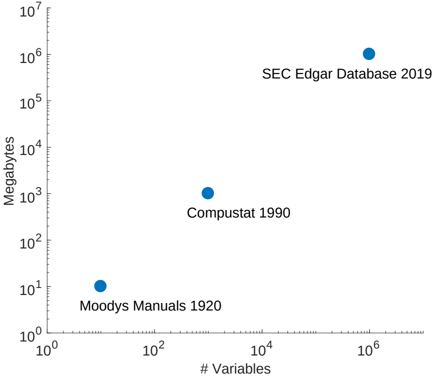
Market Efficiency in the Age of Big Data

Ian Martin   Stefan Nagel

June, 2021

Investors' Big Data problem

- Investors face huge number of potential predictors
- $\mathbb{E}\left[\text{return}\right] = f(\text{predictors})$ unknown: high-dimensional learning problem



High-dimensional learning in asset pricing

- Standard approaches in asset pricing and market efficiency testing assume rational expectations (RE)
  - Assumes away learning problem: investors know $f(\cdot)$ in
    $$\mathbb{E}[\text{return}] = f(\text{predictors})$$
  - Motivates in-sample (IS) tests of "market efficiency": IS return predictability = risk premium/mispricing
- We show: when investors learn about $f(\cdot)$ in big data setting, equilibrium asset prices exhibit in-sample predictability
- Combination of learning and big data provides clear motivation for (pseudo-)OOS testing which is lacking in RE framework

Market efficiency

- Fama (1970): A market is efficient if "prices fully reflect all available information"
- Joint hypothesis problem: your "risk premium" is my "pricing inefficiency"
- We make things simple by considering a risk-neutral world
- Then the joint hypothesis problem goes away. . . but standard tests of market efficiency break down even so

Roadmap

Two steps:

1. Investors learn about parameters of cash flow generating model and price assets accordingly
2. Econometrician analyzes equilibrium prices ex post using standard return predictability tests
   - Properties of IS tests
   - Properties of OOS tests

Setup

- N assets, $N \times J$ scaled characteristics arranged into a matrix $\boldsymbol{X}$, eg,

$$\begin{pmatrix} \text{Size}_{\text{AAPL}} & \text{Leverage}_{\text{AAPL}} & \text{Liquidity}_{\text{AAPL}} & \cdots & \text{CharJ}_{\text{AAPL}} \\ \text{Size}_{\text{AMZN}} & \text{Leverage}_{\text{AMZN}} & \text{Liquidity}_{\text{AMZN}} & \cdots & \text{CharJ}_{\text{AMZN}} \\ \text{Size}_{\text{FB}} & \text{Leverage}_{\text{FB}} & \text{Liquidity}_{\text{FB}} & \cdots & \text{CharJ}_{\text{FB}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Size}_{N} & \text{Leverage}_{N} & \text{Liquidity}_{N} & \cdots & \text{CharJ}_{N} \end{pmatrix}$$

- Normally we think of many assets and a limited number of characteristics $\implies$ fixed-J, large-N asymptotics

- This paper: many assets and many characteristics $\implies$ large-N *and* large-J asymptotics

Setup

- Investors are homogeneous and risk-neutral; interest rate is zero
- Dividend growth $\Delta \boldsymbol{y}_{t+1}$ is predictable based on characteristics $\boldsymbol{X}$:

$$\Delta \boldsymbol{y}_{t+1} = \boldsymbol{X}\boldsymbol{g} + \boldsymbol{e}_{t+1}$$

  with normalization $\frac{1}{NJ} \operatorname{tr} \boldsymbol{X}'\boldsymbol{X} = 1$

- $\boldsymbol{e}_{t+1} \sim N(\mathbf{0}, \boldsymbol{I})$ is totally unpredictable
- Dividend strips: $\boldsymbol{p}_t$ = prices at t of claims to $\boldsymbol{y}_{t+1}$
  - Think: one period $\approx$ one decade
- These assumptions are chosen to make life as simple as possible

Setup

- Prices equal expected dividends

$$\boldsymbol{p}_t = \tilde{\mathbb{E}}_t \boldsymbol{y}_{t+1} = \boldsymbol{y}_t + \tilde{\mathbb{E}}_t \Delta \boldsymbol{y}_{t+1} = \boldsymbol{y}_t + \tilde{\mathbb{E}}_t \left(\boldsymbol{X}\boldsymbol{g} + \boldsymbol{e}_{t+1} \right)$$

- Covers a range of possible assumptions about expectations $\tilde{\mathbb{E}}_t$
- Benchmarks to keep in mind...
  - Rational expectations: investors know $\boldsymbol{g}$
  - OLS: regress past cashflow growth on $\boldsymbol{X}$ to estimate $\boldsymbol{g}$
  - Random walk: give up on forecasting
  - Bayesian learning

Rational expectations: investors know $\boldsymbol{g}$

- So $\tilde{\mathbb{E}}_t \left(\boldsymbol{X}\boldsymbol{g} + \boldsymbol{e}_{t+1} \right) = \boldsymbol{X}\boldsymbol{g}$ and $\boldsymbol{p}_t = \boldsymbol{y}_t + \boldsymbol{X}\boldsymbol{g}$
- Realized returns $\boldsymbol{r}_{t+1} = \boldsymbol{y}_{t+1} - \boldsymbol{p}_t = \boldsymbol{e}_{t+1}$
- This is the usual null hypothesis that underlies market efficiency tests, orthogonality conditions, Euler equations
- But it is implausible that investors know $\boldsymbol{g}$, especially if J is large
- We focus on the case where investors must learn $\boldsymbol{g}$

Bayesian pricing framework: Prior beliefs

- Before seeing data, investors have prior beliefs

$$\boldsymbol{g} \sim N\left(\mathbf{0}, \frac{\theta}{J}\boldsymbol{I}\right), \qquad \theta > 0$$

  - Proportionality of prior covariance matrix to $\boldsymbol{I}$: can always rotate and rescale $\boldsymbol{X}$ to make it hold
  - Variance of the elements of $\boldsymbol{g}$ decline with J: ensures that variance of predictable cash flow growth does not explode when $N, J \to \infty$

- Investors then learn about $\boldsymbol{g}$ by observing $\boldsymbol{X}$ and history $\{\Delta \boldsymbol{y}_s\}_{s=1}^{t}$, summarized by sample average $\overline{\Delta \boldsymbol{y}}_t$

- As investors are risk-neutral, only the posterior mean matters

Bayesian pricing framework: Posterior mean

- Posterior mean is a ridge regression estimator

$$\tilde{\boldsymbol{g}}_t = \boldsymbol{\Gamma}_t(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\overline{\Delta\boldsymbol{y}}_t$$

i.e., OLS estimator shrunk towards prior mean by the matrix

$$\boldsymbol{\Gamma}_t = \boldsymbol{Q}\left(\boldsymbol{I} + \frac{J}{N\theta t}\boldsymbol{\Lambda}^{-1}\right)^{-1}\boldsymbol{Q}'$$

where $\boldsymbol{Q}$, $\boldsymbol{\Lambda}$ come from PC decomposition $\frac{1}{N}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'$

Bayesian pricing framework: Posterior mean

- Posterior mean is a ridge regression estimator

$$\tilde{\boldsymbol{g}}_t = \boldsymbol{\Gamma}_t(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\overline{\Delta\boldsymbol{y}}_t$$

  i.e., OLS estimator shrunk towards prior mean by the matrix

$$\boldsymbol{\Gamma}_t = \boldsymbol{Q}\left(\boldsymbol{I} + \frac{J}{N\theta t}\boldsymbol{\Lambda}^{-1}\right)^{-1}\boldsymbol{Q}'$$

  where $\boldsymbol{Q}$, $\boldsymbol{\Lambda}$ come from PC decomposition $\frac{1}{N}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'$

- Shrinkage strong
  - if t small (short time dimension)

Bayesian pricing framework: Posterior mean

- Posterior mean is a ridge regression estimator

$$\tilde{\boldsymbol{g}}_t = \boldsymbol{\Gamma}_t(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\overline{\Delta \boldsymbol{y}}_t$$

i.e., OLS estimator shrunk towards prior mean by the matrix

$$\boldsymbol{\Gamma}_t = \boldsymbol{Q}\left(\boldsymbol{I} + \frac{J}{N\theta t}\boldsymbol{\Lambda}^{-1}\right)^{-1}\boldsymbol{Q}'$$

where $\boldsymbol{Q}$, $\boldsymbol{\Lambda}$ come from PC decomposition $\frac{1}{N}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'$

- Shrinkage strong
  - if t small (short time dimension)
  - if θ small (prior tightly concentrated around zero)

Bayesian pricing framework: Posterior mean

- Posterior mean is a ridge regression estimator

$$\tilde{\boldsymbol{g}}_t = \boldsymbol{\Gamma}_t(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\overline{\Delta\boldsymbol{y}}_t$$

i.e., OLS estimator shrunk towards prior mean by the matrix

$$\boldsymbol{\Gamma}_t = \boldsymbol{Q}\left(\boldsymbol{I} + \frac{J}{N\theta t}\boldsymbol{\Lambda}^{-1}\right)^{-1}\boldsymbol{Q}'$$

where $\boldsymbol{Q}$, $\boldsymbol{\Lambda}$ come from PC decomposition $\frac{1}{N}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'$

- Shrinkage strong
  - if t small (short time dimension)
  - if θ small (prior tightly concentrated around zero)
  - if J/N is large (many predictors)

Bayesian pricing framework: Posterior mean

- Posterior mean is a ridge regression estimator

$$\tilde{\boldsymbol{g}}_t = \boldsymbol{\Gamma}_t (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\overline{\Delta \boldsymbol{y}}_t$$

  i.e., OLS estimator shrunk towards prior mean by the matrix

$$\boldsymbol{\Gamma}_t = \boldsymbol{Q}\left(\boldsymbol{I} + \frac{J}{N\theta t}\boldsymbol{\Lambda}^{-1}\right)^{-1}\boldsymbol{Q}'$$

  where $\boldsymbol{Q}$, $\boldsymbol{\Lambda}$ come from PC decomposition $\frac{1}{N}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'$

- Shrinkage strong
  - if t small (short time dimension)
  - if θ small (prior tightly concentrated around zero)
  - if J/N is large (many predictors)
  - along unimportant principal components of $\boldsymbol{X}$ (small eigenvalues)

Equilibrium realized returns

Proposition

With assets priced based on $\tilde{\boldsymbol{g}}_t$, realized returns are

$$\boldsymbol{r}_{t+1} = \boldsymbol{y}_{t+1} - \boldsymbol{p}_t = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{\Gamma}_t)\boldsymbol{g} - \boldsymbol{X}\boldsymbol{\Gamma}_t(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\bar{\boldsymbol{e}}_t + \boldsymbol{e}_{t+1}$$

where $\bar{\boldsymbol{e}}_t = \frac{1}{t}\sum_{s=1}^{t} \boldsymbol{e}_s$

Equilibrium realized returns

Proposition

With assets priced based on $\tilde{\boldsymbol{g}}_t$, realized returns are

$$\boldsymbol{r}_{t+1} = \boldsymbol{y}_{t+1} - \boldsymbol{p}_t = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{\Gamma}_t)\boldsymbol{g} - \boldsymbol{X}\boldsymbol{\Gamma}_t(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\bar{\boldsymbol{e}}_t + \boldsymbol{e}_{t+1}$$

where $\bar{\boldsymbol{e}}_t = \frac{1}{t}\sum_{s=1}^{t} \boldsymbol{e}_s$

- "underreaction" to $\boldsymbol{X}$ due to shrinkage

Equilibrium realized returns

Proposition

With assets priced based on $\tilde{\boldsymbol{g}}_t$, realized returns are

$$\boldsymbol{r}_{t+1} = \boldsymbol{y}_{t+1} - \boldsymbol{p}_t = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{\Gamma}_t)\boldsymbol{g} - \boldsymbol{X}\boldsymbol{\Gamma}_t(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\bar{\boldsymbol{e}}_t + \boldsymbol{e}_{t+1}$$

where $\bar{\boldsymbol{e}}_t = \frac{1}{t}\sum_{s=1}^{t} \boldsymbol{e}_s$

- "underreaction" to $\boldsymbol{X}$ due to shrinkage
- "overreaction" to estimation error in $\tilde{\boldsymbol{g}}_t$, dampened by shrinkage

Equilibrium realized returns

Proposition

With assets priced based on $\tilde{\boldsymbol{g}}_t$, realized returns are

$$\boldsymbol{r}_{t+1} = \boldsymbol{y}_{t+1} - \boldsymbol{p}_t = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{\Gamma}_t)\boldsymbol{g} - \boldsymbol{X}\boldsymbol{\Gamma}_t(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\bar{\boldsymbol{e}}_t + \boldsymbol{e}_{t+1}$$

where $\bar{\boldsymbol{e}}_t = \frac{1}{t}\sum_{s=1}^{t} \boldsymbol{e}_s$

- "underreaction" to $\boldsymbol{X}$ due to shrinkage
- "overreaction" to estimation error in $\tilde{\boldsymbol{g}}_t$, dampened by shrinkage
- unpredictable shock (the only term in RE case)

Predictive coefficient estimates, $\boldsymbol{h}_{t+1}$

- Econometrician cross-sectionally regresses (OLS)

$$\boldsymbol{r}_{t+1} = \underbrace{\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{\Gamma}_t)\boldsymbol{g}}_{\text{“underreaction”}} - \underbrace{\boldsymbol{X}\boldsymbol{\Gamma}_t(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\bar{\boldsymbol{e}}_t}_{\text{“overreaction”}} + \underbrace{\boldsymbol{e}_{t+1}}_{\text{RE}}$$

  on characteristics matrix $\boldsymbol{X}$ and obtains predictive coefficients

$$\boldsymbol{h}_{t+1} = (\boldsymbol{I} - \boldsymbol{\Gamma}_t)\boldsymbol{g} - \boldsymbol{\Gamma}_t \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\bar{\boldsymbol{e}}_t + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{e}_{t+1}$$

- “Kitchen sink” regression approximates what many individual studies have done collectively (“factor zoo”)

In-sample predictability test: RE null

- Consider the return predictability test statistic

$$T_{re} \equiv \frac{\boldsymbol{h}'_{t+1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{h}_{t+1} - J}{\sqrt{2J}}$$

- Standard approach takes RE as null hypothesis, which implies

$$\boldsymbol{h}_{t+1} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{e}_{t+1}$$

- If so,

$$T_{re} \xrightarrow{d} N(0,1) \quad \text{as } N, J \to \infty,\ J/N \to \psi > 0$$

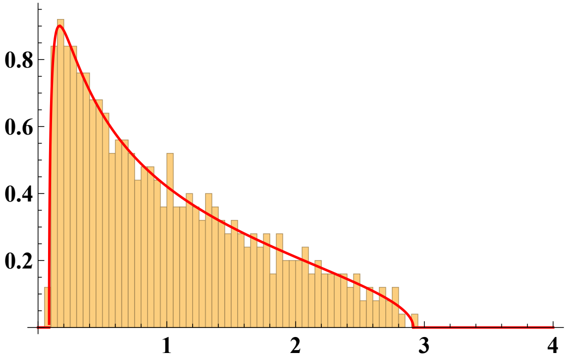
Big data

What happens as $J, N \to \infty$ with $J/N \to \psi > 0$?

- Case 1: A few principal components summarize the data
  - Formally: the eigenvalues of $\frac{1}{N}\boldsymbol{X}'\boldsymbol{X}$ tend to zero
  - Then market efficiency test works as usual, $T_{re} \xrightarrow{d} N(0,1)$
- Case 2: "Big data"
  - Formally: the eigenvalues of $\frac{1}{N}\boldsymbol{X}'\boldsymbol{X}$ are $> \varepsilon$
  - Happens if, eg, the entries of $\boldsymbol{X}$ are iid random variables
- We are interested in Case 2

Eigenvalues in an example with iid random $\boldsymbol{X}$

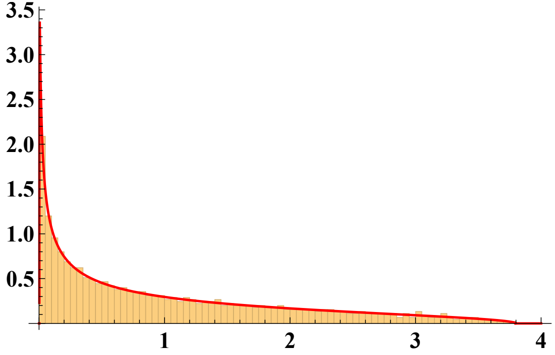
Histogram for $N = 1000$, $J = 500$ & asymptotic limit as $N, J \to \infty$ with $\frac{J}{N} = 0.5$



- When J is large, random data matrix $\boldsymbol{X}$ has many columns that are roughly orthogonal, hence many eigenvalues close to zero
- But our big data assumption is satisfied (Bai and Yin, 1993)

Eigenvalues in an example with iid random $\boldsymbol{X}$

Histogram for $N = 1000$, $J = 900$ & asymptotic limit as $N, J \to \infty$ with $\frac{J}{N} = 0.9$



- When J is large, random data matrix $\boldsymbol{X}$ has many columns that are roughly orthogonal, hence many eigenvalues close to zero
- But our big data assumption is satisfied (Bai and Yin, 1993)

In-sample predictability test with big data

Proposition

In equilibrium, the test statistic T_{re} satisfies

$$\frac{T_{re}}{\sqrt{\mu^2+\sigma^2}} - \frac{\mu-1}{\sqrt{2\left(\mu^2+\sigma^2\right)}}\sqrt{J} \xrightarrow{d} N(0,1)$$

where $1<\mu<2$ and $1<\sqrt{\mu^2+\sigma^2}<2$ are determined by eigenvalues

- Therefore,

$$T_{re} \approx \sqrt{\mu^2+\sigma^2}\, N(0,1) + \frac{\mu-1}{\sqrt{2}}\sqrt{J}$$

- In a big data world, we are almost certain to reject the RE null

Interpretation as a trading strategy

- Consider a characteristics-based trading strategy with weights

$$\boldsymbol{w}_{IS,t} = \boldsymbol{X}\boldsymbol{h}_{t+1}, \qquad r_{IS,t+1} = \boldsymbol{w}'_{IS,t}\boldsymbol{r}_{t+1}$$

("in-sample" because $\boldsymbol{h}_{t+1}$ estimated using returns $\boldsymbol{r}_{t+1}$)

- We can rewrite $r_{IS,t+1} = \boldsymbol{h}'_{t+1}\boldsymbol{X}'\boldsymbol{r}_{t+1}$ as

$$r_{IS,t+1} = \boldsymbol{h}'_{t+1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{h}_{t+1}$$

- Econometrician's test is equivalent to checking whether the trading strategy does well

Conclusions so far

- Asset returns under high-dimensional learning are very different from asset returns under RE, or in a "small data" world
- IS return predictability need not be consequence of risk premia or behavioral biases
- Not an econometric issue: the RE null is simply false, because learning + big data makes returns predictable in sample even without risk premia or behavioral biases
- Existence of a "factor zoo" based on IS predictability evidence not surprising in high-dimensional setting

(Absence of) out-of-sample return predictability

Proposition

Consider an out-of-sample strategy with predicted returns as portfolio weights, $r_{OOS,t+1} = \boldsymbol{r}'_{t+1}\boldsymbol{X}\boldsymbol{h}_{s+1}$ *where* $t \neq s$*. Then* $\mathbb{E}\, r_{OOS,t+1} = 0$

- Forward case $t > s$ is natural: Investors are Bayesian so the econometrician cannot "beat" investors
- Backward case $t < s$ is more surprising. Not a tradable strategy, but interesting for research
  - Suggests backwards OOS tests (e.g., Linnainmaa and Roberts 2018) and cross-validation (e.g., Kozak, Nagel and Santosh 2020; Bryzgalova, Pelger, and Zhu 2020) could be appropriate for Bayesian learning setting

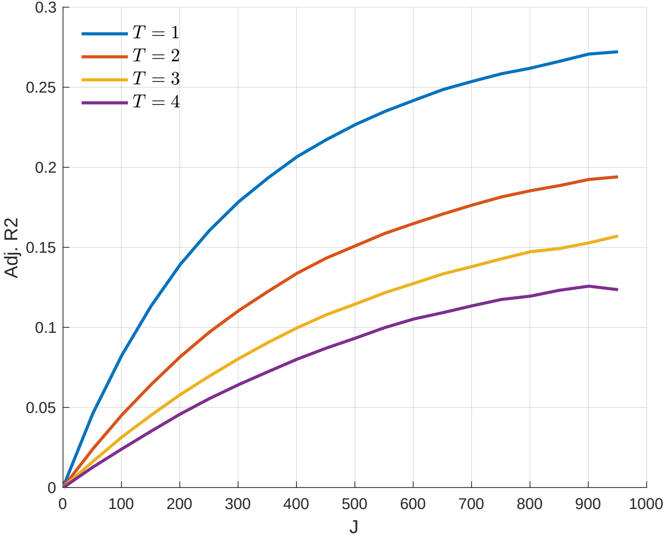
Finite-sample analysis: Simulations

- Simulate cash-flows, prices, returns for $N = 1000$ assets
- To generate data, we set $\theta = 1$ in

$$\boldsymbol{\Sigma}_g = \frac{\theta}{J}\boldsymbol{I}$$

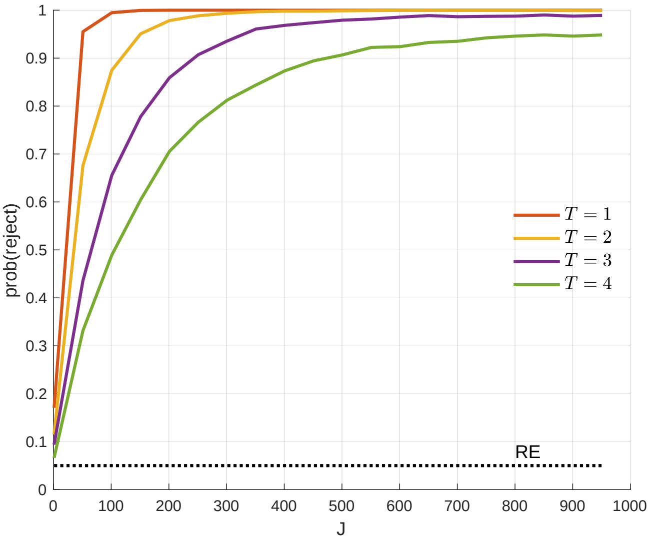
- θ = ratio of forecastable/residual cash-flow growth variance
  - Based on analyst expectations, Chen, Karceski, and Lakonishok (2003) find forecastable/residual cash-flow growth variance of 0.4 at 10yr horizon
- Econometrician regresses $\boldsymbol{r}_{T+1}$ on $\boldsymbol{X}$ after investors have learned about $\boldsymbol{g}$ for T periods

Adjusted R^2



0.3
0.25
0.2
0.15
0.1
0.05
0
Adj. R2
0
100
200
300
400
500
600
700
800
900
1000
J
T = 1
T = 2
T = 3
T = 4

Rejection probability of no-return-predictability null



Variations in the paper

- So far, shrinkage was purely due to objectively correct informative prior beliefs of investors
- If (time-varying?) cost to observe predictor variables, this may induce excess shrinkage $\implies$ positive OOS returns
- Similar results when investors deal with big data by using Lasso rather than ridge regressions

Empirical illustration: IS vs OOS predictability

- Suppose returns from earlier are augmented with a risk premium/mispricing component $\boldsymbol{X\gamma}$

$$\boldsymbol{r}_{t+1} = \color{red}{\boldsymbol{X\gamma}} + \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{\Gamma}_t)\boldsymbol{g} - \boldsymbol{X\Gamma}_t(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\bar{\boldsymbol{e}}_t + \boldsymbol{e}_{t+1}$$

- The presence of the over/underreaction terms on the RHS will bias OLS estimates of $\boldsymbol{X\gamma}$

- But OOS returns measure importance of risk premium/mispricing:

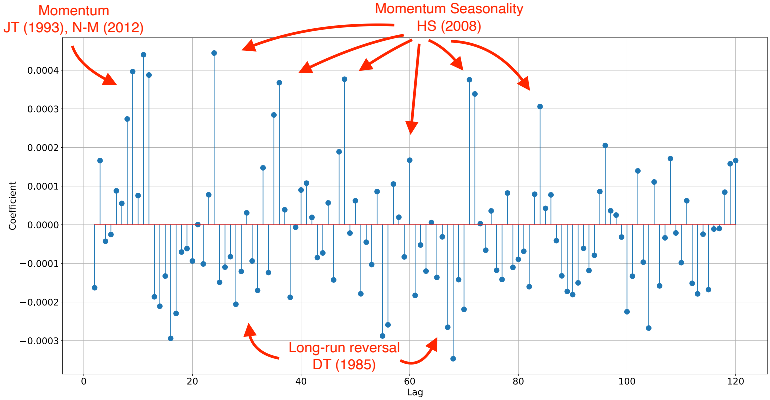
$$\frac{1}{N}\boldsymbol{\gamma}'\boldsymbol{X}'\boldsymbol{X\gamma} = \mathbb{E}\left[r_{OOS,t+1}\right]$$

Empirical illustration: IS vs OOS predictability

- To illustrate, we seek a large set of predictors that have been available to investors over a long period
- Use past returns of each stock (available, in principle, for decades) to predict returns in month t with
  - Returns in months $t-2, ..., t-120$
  - Squared returns in months $t-2, ..., t-120$
- All U.S. stocks on CRSP, except market cap $<$ 20th NYSE percentile or price $<$ \$1 at the end of month $t-1$
- All predictors cross-sectionally demeaned and standardized to unit S.D. each month
- Ridge regression with leave-one-year-out cross-validation to choose penalty parameter value

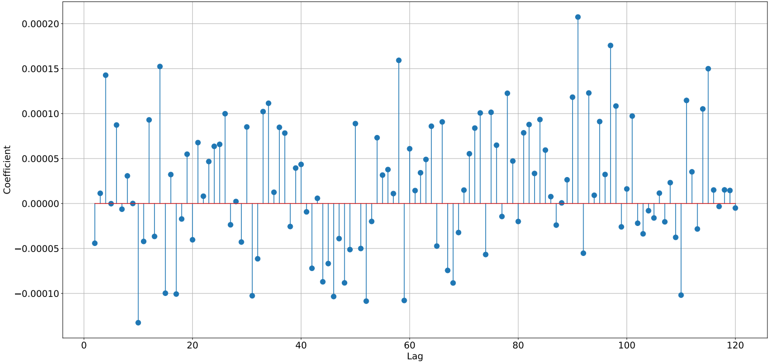
In-sample: Past return coefficients

Sample period: 1971-2018



In-sample: Past squared return coefficients

Sample period: 1971-2018



Estimating risk premia/mispricing in presence of learning

- Recall: Estimate $\gamma' \boldsymbol{X}' \boldsymbol{X} \gamma$ from sample version of

$$\gamma' \boldsymbol{X}' \boldsymbol{X} \gamma = \mathbb{E}\left[r_{OOS,t+1}\right]$$

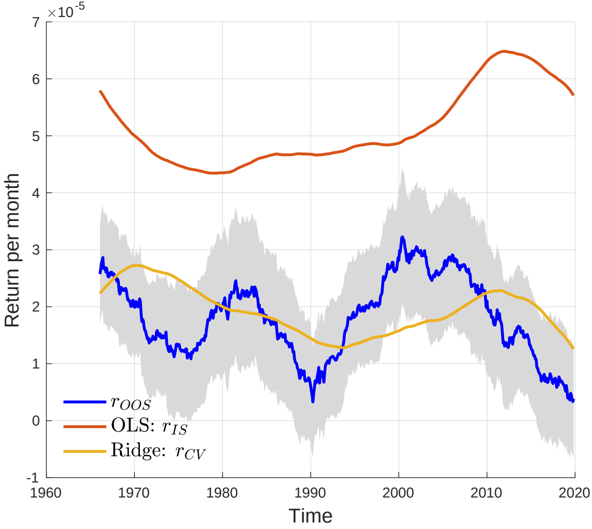
- $\boldsymbol{h}_t$ estimated with OLS in backwards 20-year moving window up to month t and used to form

$$r_{OOS,t+1} = \boldsymbol{r}'_{t+1} \boldsymbol{X} \boldsymbol{h}_t$$

- $r_{OOS,t+1}$ averaged in 10-year moving windows

- Compare with two other returns in backwards 20-year window
  - In-sample return based on OLS estimates r_{IS}
  - Return on validation folds for cross-validated ridge regression r_{CV}

Estimating risk premia/mispricing in presence of learning



IS vs OOS returns

- In an RE model, expected IS and OOS portfolio returns would both equal $\gamma' X' X \gamma$
- If investors learn, this is still true for the OOS portfolio return
- But the IS return is distorted by learning-induced components that are not predictable OOS
- Seems that the learning case is relevant
- IS predictability does not carry over to OOS predictability and hence does not reflect risk premia demanded by investors ex ante, or persistent belief distortions

Implications: Market Efficiency in the Age of Big Data

- In Big Data setting, RE (investors know $\boldsymbol{g}$) is implausible
- Learning (about $\boldsymbol{g}$) has strong effects on asset prices
- Risk premia & bias theories should focus on explaining OOS, not IS, return predictability
- Investor learning provides clear motivation for (pseudo-)OOS testing which is lacking in RE framework
- Broader question: Can puzzling phenomena be explained by agents learning in Big Data settings?