Empirical Asset Pricing

Return predictability

Jonas Nygaard Eriksen

Department of Economics and Business Economics
Aarhus University and CREATES

Spring 2022

Outline

1. Introduction
2. In-sample predictability
   - Stambaugh bias
   - A parametric bootstrap
   - Long-horizon predictability
3. Out-of-sample predictability
   - Statistical evaluation
   - Economic evaluation
   - Time-varying predictability
4. Other asset classes
5. Potential projects

Return predictability in a nutshell

The fundamental question

An ongoing and long-standing debate in finance is whether excess returns (risk premia) on financial assets are predictable

- 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

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

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 $\boldsymbol{x}_t$

$$\mathbb{E}_t\left[r_{t+1}\right] = \widehat{\alpha} + \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}} \tag{2}$$

- 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

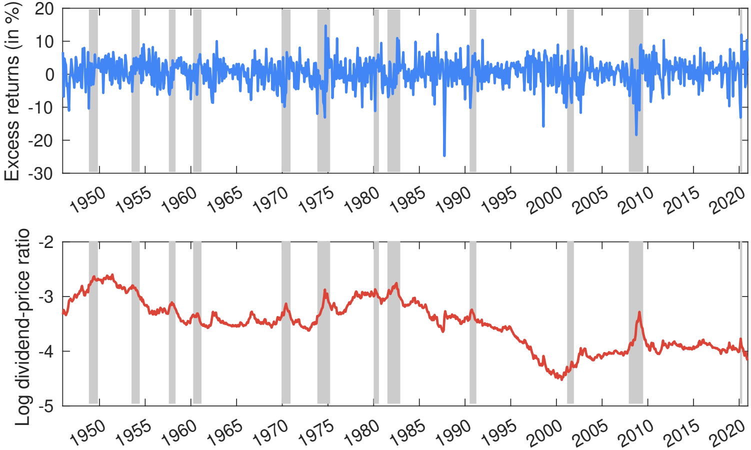
  1. Can help help answer a long-standing debate: Does expected excess returns (risk premia) vary over time?
  2. Can help sharpen the distinction between risk-based (market efficiency) and behavioral (ineffiency) explanations for variations in returns
  3. Results can lead to better and more realistic asset pricing models
  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

S&P500 excess returns and dividend-price ratio

* We will consider a classic example of stock return predictability using 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

$$
\begin{aligned}
r_{t+1} &= \ln\left(P_{t+1} + D_{t+1}\right) - \ln\left(P_t\right) && (3) \\
&= p_{t+1} - p_t + \ln\left(1 + \exp\left(d_{t+1} - p_{t+1}\right)\right) && (4)
\end{aligned}
$$

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

$$
r_{t+1} \approx \kappa + \phi p_{t+1} + (1-\phi)\, d_{t+1} - p_t \tag{5}
$$

with

$$
\phi = \frac{1}{1 + \exp\left(\overline{d - p}\right)} \tag{6}
$$

and

$$
\kappa = -\ln\phi - (1-\phi)\ln\left(1/\phi - 1\right) \tag{7}
$$

Why the dividend-price ratio is a natural predictor

* Solving (5) 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

$$p_t = \frac{\kappa}{1-\phi} + \mathbb{E}_t \left[\sum_{j=0}^{\infty} \phi^j \left[(1-\phi)\, d_{t+1+j} - r_{t+1+j} \right] \right] \tag{8}$$

* Similarly, (5) can be stated in terms of the log dividend-price ratio and iterating forward as above yields

$$d_t - p_t = \frac{-\kappa}{1-\phi} + \mathbb{E}_t \left[\sum_{j=0}^{\infty} \phi^j \left[-\Delta d_{t+1+j} + r_{t+1+j} \right] \right] \tag{9}$$

  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

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)

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim iid\left(0,\sigma_\varepsilon^2\right) \tag{10}$$

$$x_{t+1} = \lambda + \rho x_t + \nu_{t+1}, \quad \nu_{t+1} \sim iid\left(0,\sigma_\nu^2\right) \tag{11}$$

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

- The OLS estimate $\widehat{\rho}$ in (11) is downward-biased in a finite sample (Kendall, 1954)

$$\mathbb{E}\left[\widehat{\rho} - \rho\right] \approx -\left(\frac{1+3\rho}{T}\right) \tag{12}$$

- Stambaugh (1986, 1999) shows that the bias in $\widehat{\beta}$ in (10) is then given by

$$\mathbb{E}\left[\widehat{\beta} - \beta\right] = \underbrace{\frac{\sigma_{\varepsilon\nu}}{\sigma_\nu^2}}_{\gamma} \mathbb{E}\left[\widehat{\rho} - \rho\right] \tag{13}$$

Interpreting the bias

- Note that we can interpret the term γ in (13) as a regression coefficient

$$\varepsilon_{t+1} = \gamma \nu_{t+1} + \eta_{t+1} \tag{14}$$

- Given (13) and (12), the bias in $\widehat{\beta}$ can be approximated as

$$\mathbb{E}\left[\widehat{\beta} - \beta\right] \approx -\frac{\sigma_{\varepsilon\nu}}{\sigma_\nu^2}\left(\frac{1+3\rho}{T}\right) = -\gamma\left(\frac{1+3\rho}{T}\right) \tag{15}$$

- 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 $(d_t - pt)$, 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}

- 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

$$\begin{bmatrix} \varepsilon_{t+1} \\ \nu_{t+1} \end{bmatrix} \sim iid\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma\right), \quad \text{where} \quad \Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\nu} \\ \sigma_{\nu\varepsilon} & \sigma_\nu^2 \end{bmatrix} \tag{16}$$

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

$$x^*_1 \sim \mathcal{N}\left(\overline{x},\widehat{\sigma}^2_x\right) \tag{17}$$

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

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

$$\widehat{\lambda} + \widehat{\rho}x^*_t + \nu^*_{t+1} \to x^*_{t+1} \tag{18}$$

Parametric bootstrap example

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}

$$\widehat{\alpha} + \beta_0 x_t^* + \varepsilon_{t+1}^* \to r_{t+1}^* \tag{19}$$

6. Estimate the predictive regression in (10) on the simulated data to obtain $\widetilde{\beta}^{(1)}$

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

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

$$\mathbb{P}\left[\widetilde{\beta} > \widehat{\beta}\right] = \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}\left\{\widetilde{\beta}^{(i)} > \widehat{\beta}\right\} \tag{20}$$

* 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

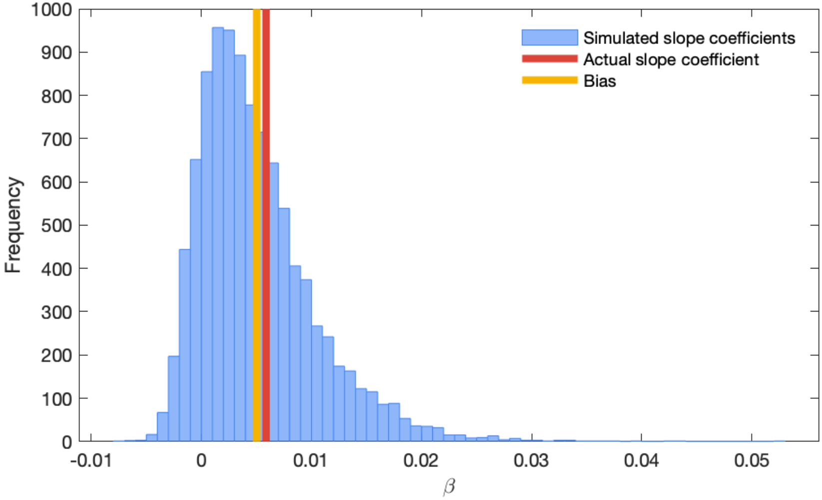
$$\text{Bias}\left(\widehat{\beta}\right) = \frac{1}{M}\sum_{i=1}^{M}\widetilde{\beta}^{(i)} - \beta_0 \tag{21}$$

* 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 (10) and (11) 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

$$\mathbb{E}\left[\widehat{\beta} - \beta | \widehat{\rho}, \rho\right] = \gamma \left[\widehat{\rho} - \rho\right] \tag{22}$$

- 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

$$\widehat{\beta}_{\text{Adj.}} = \widehat{\beta} - \gamma \left[\widehat{\rho} - 1\right] \tag{23}$$

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

$$\text{Var}\left[\widehat{\beta}_{\text{Adj.}}\right] = \frac{\sigma_\eta^2}{(T\sigma_x^2)} \tag{24}$$

  where $\sigma_x^2 = T^{-1} \sum_{t=1}^{T} \left(x_t - \overline{x}_t\right)^2$ is the (biased) sample variance of x_t and σ_η^2 is the variance of the residuals from (14)

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

$$\Delta d_{t+1} = \alpha_d + \beta_d (d_t - p_t) + \varepsilon_{d,t+1} \tag{25}$$

$$r_{t+1} = \alpha_r + \beta_r (d_t - p_t) + \varepsilon_{r,t+1} \tag{26}$$

$$d_{t+1} - p_{t+1} = \lambda + \rho (d_t - p_t) + \varepsilon_{dp,t+1} \tag{27}$$

- The loglinear approximation in (5) implies that the parameters are intimately linked

$$\beta_r = 1 + \beta_d - \phi\rho \tag{28}$$

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

$$\beta_r > \beta_d + 0.03 \tag{29}$$

- 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)

Long-horizon predictability

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 (1)

$$r_{t \to t+h} = \alpha_h + \beta_h x_t + \varepsilon_{t \to t+h} \tag{30}$$

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

$$r_{t \to t+h} = \ln\left(1 + r_{t \to t+h}\right) = \sum_{i=1}^{h} \ln\left(1 + r_{t+i}\right) \tag{31}$$

$$= \sum_{i=1}^{h} r_{t+i} \tag{32}$$

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 (10) and (11) without constant terms for notational ease, then for $h = 2$ we have

$$r_{t+1} + r_{t+2} = \beta x_t + \varepsilon_{t+1} + \beta x_{t+1} + \varepsilon_{t+2} \tag{33}$$

$$= \beta x_t + \varepsilon_{t+1} + \beta \rho x_t + \beta \nu_{t+1} + \varepsilon_{t+2} \tag{34}$$

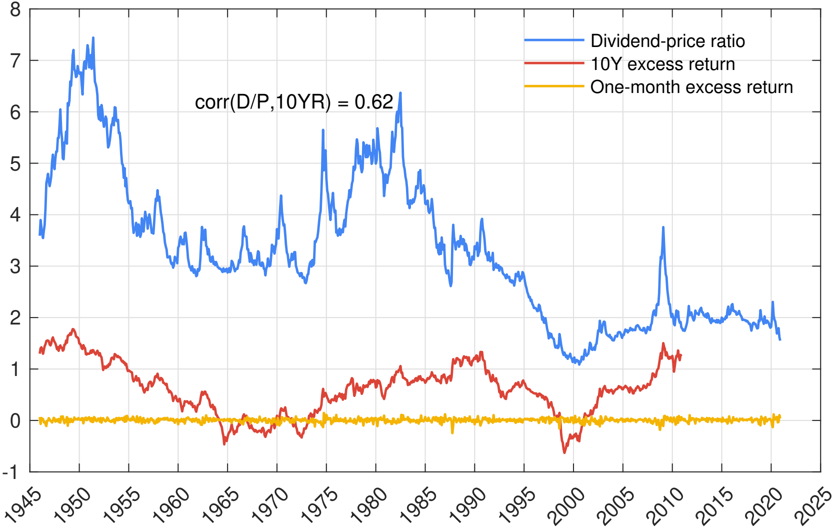
$$= \beta \left(1 + \rho\right) x_t + \beta \nu_{t+1} + \varepsilon_{t+1} + \varepsilon_{t+2} \tag{35}$$

- Similarly, for $h = 3$ we obtain

$$r_{t+1} + r_{t+2} + r_{t+3} = \beta \left(1 + \rho + \rho^2\right) x_t + \text{error terms} \tag{36}$$

- 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 (30) 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

* 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

$$\underbrace{t = 1,2,\ldots,R}_{\text{In-sample}}, \underbrace{R+1,R+2,\ldots,T}_{\text{Out-of-sample}} \tag{37}$$

* 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!)

$$\overline{r}_{t+1} = \frac{1}{t} \sum_{i=1}^{t} r_i \tag{38}$$

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

$$\text{MSFE} = \frac{1}{T-R} \sum_{i=R+1}^{T} \left(r_i - \widehat{r}_i\right)^2 \tag{39}$$

- 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)

$$\text{R}^2_{OS} = 1 - \frac{\text{MSFE}_x}{\text{MSFE}_{HA}} = 1 - \frac{\sum_{i=R+1}^{T} \left(r_i - \widehat{r}_i\right)^2}{\sum_{i=R+1}^{T} \left(r_i - \overline{r}_i\right)^2} \tag{40}$$

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

$$\mathcal{H}_0 : R^2_{OS} \leq 0 \quad \text{(No predictability)} \tag{41}$$

$$\mathcal{H}_1 : R^2_{OS} > 0 \tag{42}$$

* 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

$$d_i = (r_i - \overline{r}_i)^2 - (r_i - \widehat{r}_i)^2, \quad i = R+1, R+2, \ldots, T \tag{43}$$

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

$$d_i = \theta + \epsilon_i \tag{44}$$

  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

$$f_i = (r_i - \overline{r}_i)^2 - \left[(r_i - \widehat{r}_i)^2 - (\overline{r}_i - \widehat{r}_i)^2\right], \quad i = R+1, R+2, \ldots, T \tag{45}$$

  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

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

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

$$\text{CDSFE}_t = \sum_{i=R+1}^{t} (r_i - \overline{r}_i)^2 - \sum_{i=R+1}^{t} (r_i - \widehat{r}_i)^2, \quad t > i \tag{47}$$

- 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

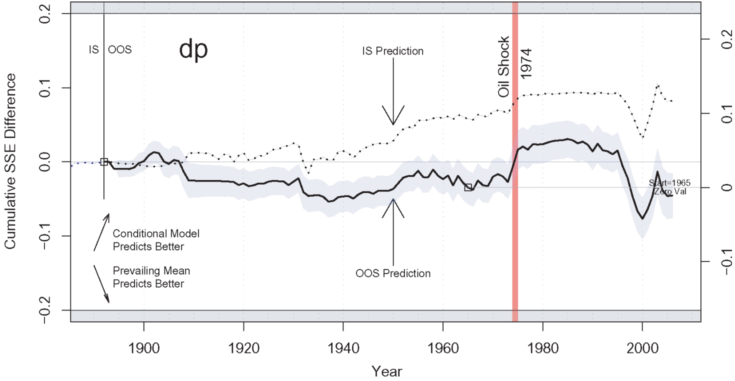
- Goyal and Welch (2008) examine a long list of potential predictors and find limited evidence of in-sample predictability and essentially no evidence supporting out-of-sample predictability

- 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	$\bar{R}^2$	OOS $\bar{R}^2$	$\bar{R}^2$	ΔRMSE	Power	OOS $\bar{R}^2$	$\bar{R}^2$	ΔRMSE	Power	$\bar{R}^2$
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

$$\widehat{r}_{t+1} = \max\left\{0, \widehat{\alpha} + \max\left\{0, \widehat{\beta}\right\} x_t\right\} \tag{48}$$

- 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

- 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

$$\max_{\omega_t} \mathbb{E}_t\left[r_{p,t+1}\right] - \frac{1}{2}\gamma \mathsf{Var}_t\left[r_{p,t+1}\right] \tag{49}$$

- 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

$$\omega_t = \left(\frac{1}{\gamma}\right)\frac{\mathbb{E}_t\left[r_{t+1} - r_{f,t+1}\right]}{\mathsf{Var}_t\left[r_{t+1} - r_{f,t+1}\right]}, \tag{50}$$

  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

$$r_{p,t+1} = (1-\omega_t)\, r_{f,t+1} + \omega_t r_{t+1} = r_{f,t+1} + \omega_t\left(r_{t+1} - r_{f,t+1}\right) \tag{51}$$

Certainty equivalent returns (utility gains)

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

$$\text{CER} = \mu_p - \frac{1}{2}\gamma\sigma_p^2 \tag{52}$$

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

$$\Delta = 1200 \times (\text{CER}_x - \text{CER}_{HA}) \tag{53}$$

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

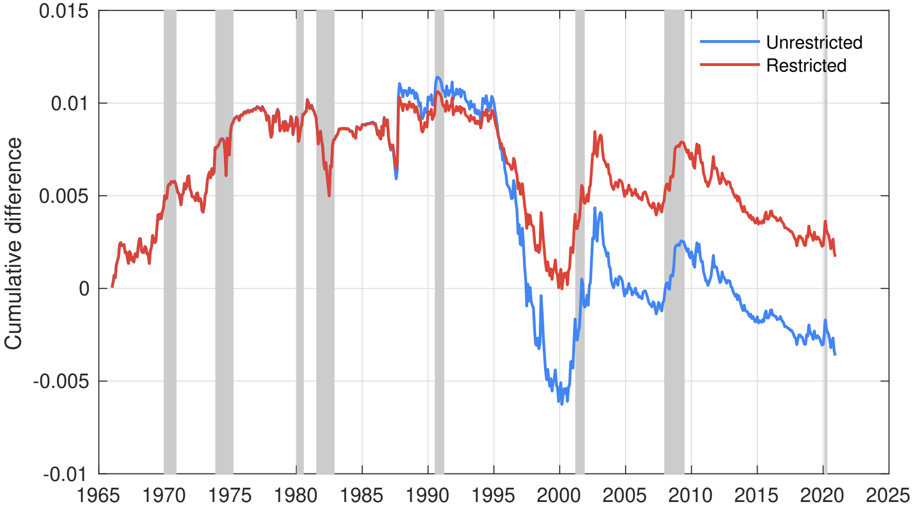
- We sometimes restrict the weights ω_t to lie between, say, $-0.5 \geq \omega_t \geq 1.5$ or similar to avoid extreme positions, i.e., we impose reasonable shorting and leverage constraints

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

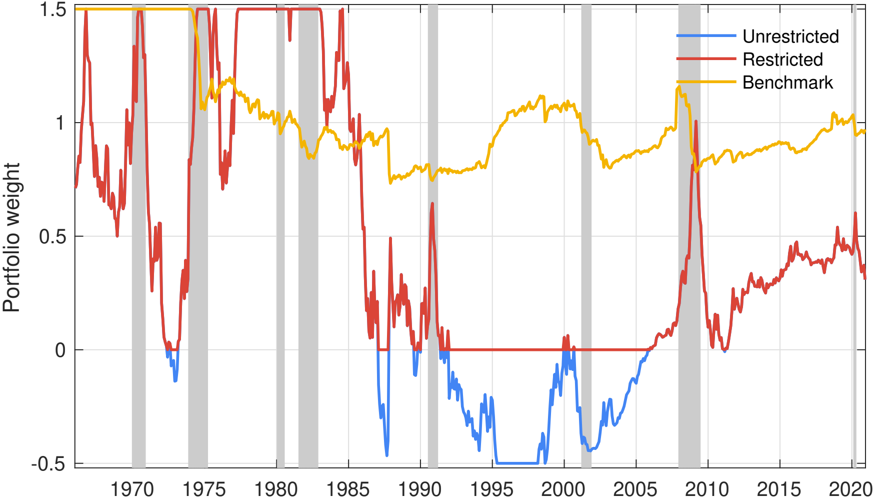
- Let us continue our example using the log dividend-price ratio to predict excess stock market returns in an out-of-sample setting
  - The initial window is $R = 240$ observations and we consider an expanding window forecasting scheme
  - The relative risk aversion is set to $\gamma = 3$ in the economic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

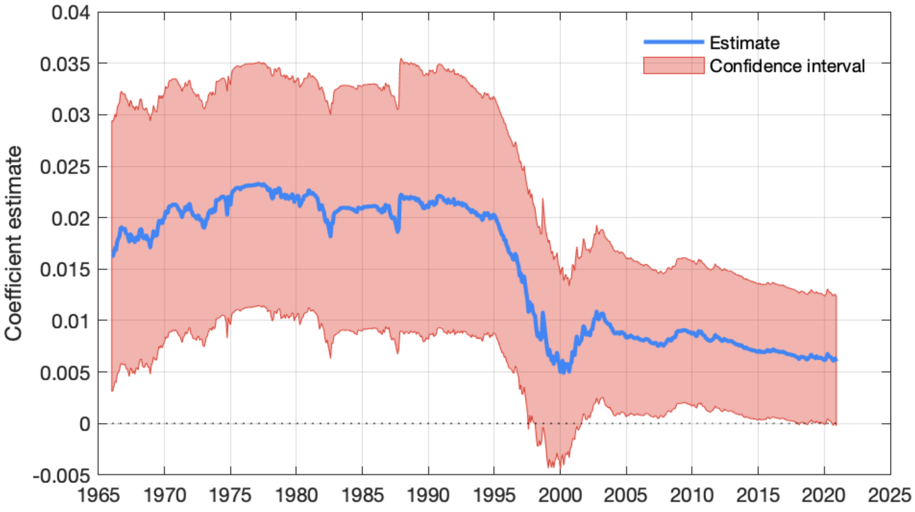
Goyal-Welch CDSFE plot



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)

$$r_{t+1} = \alpha + \beta dp_t + \varepsilon_{t+1} \tag{54}$$

* 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

$$r_{t+1} = \alpha + \boldsymbol{x}_t' \boldsymbol{\beta} + \varepsilon_{t+1} \tag{55}$$

* **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:** In FX markets, one can either consider forecasting log spot changes $\Delta s_{t+1} = s_{t+1} - s_t$ or currency excess returns $rx_{t+1} = f_t - s_{t+1} \approx i_t^{\star} - i_t - \Delta s_{t+1}$ by regressing it on various 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))

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.

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.

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.

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.

References

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.

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.

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.

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.

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

References

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.

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.

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.