Let´s Do It Again: Bagging Equity Premium Predictors

Eric Hillebrand[1], Tae-Hwy Lee[2], and Marcelo C. Medeiros[3]

[1]Louisiana State University

[2]University of California, Riverside

[3]Pontifical Catholic University of Rio de Janeiro

October 17, 2008

One-page Summary

- Restrictions in the forecast model

$$y_{t+1} = \alpha + \beta x_t + \varepsilon_t, \ t = 1, \dots, T - 1.$$

  1. $\beta > 0$.
  2. $\mathbb{E}(y_{T+1}|\mathcal{F}_T) > 0$.
- Motivation: Discussion in Campbell and Thompson (2008) and Goyal and Welch (2008) (both forthcoming RFS).
- Apply Gordon and Hall (2008): Bagging approach to imposing restrictions on parameters. We extend to forecast situation (contribution 1).
- Using Campbell and Thompson (2008) data set, find forecast improvements using this bagging approach (contribution 2).

Restrictions

- Considering the forecast equation

$$y_{t+1} = \alpha + \beta x_t + \varepsilon_t, \; t = 1, \ldots, T-1,$$

we define the forecasts:

- Unrestricted (**UF**) Goyal and Welch (2008)

$$\tilde{y}_{T+1} = \tilde{\alpha} + \tilde{\beta} x_T,$$

$\tilde{\alpha}, \tilde{\beta}$ unrestricted OLS estimators.

- Positive coefficient restriction (**PC**) Campbell and Thompson (2008)

$$\bar{y}_{T+1} = \bar{\alpha} + \bar{\beta} x_T.$$

$\bar{\beta} = \max\{\tilde{\beta}, 0\}, \; \bar{\alpha} = \mathbb{1}_{\{\tilde{\beta} \geq 0\}} \tilde{\alpha} + \mathbb{1}_{\{\tilde{\beta} < 0\}} \sum_{t=1}^{T} y_t / T.$

Gordon and Hall (2008) bagging estimator

- Gordon and Hall (2008) estimator of parameter θ.

$$\hat{\theta} = \frac{1}{B}\sum_{j=1}^{B} \max\{\tilde{\theta}^{*(j)}, 0\} = \frac{1}{B}\sum_{j=1}^{B} \bar{\theta}^{*(j)}$$

Restrictions 2

- Gordon-Hall forecast (**PC-GH**)

$$\hat{y}_{T+1} = \hat{\alpha} + \hat{\beta} x_T$$

- Positive forecast (**PF**)

$$y_{T+1}^{PF} = \mathbb{1}_{\{\tilde{y}_{T+1} \geq 0\}} \tilde{y}_{T+1}$$

- Positive coefficient and positive forecast (**PCF**)

$$y_{T+1}^{PCF} = \mathbb{1}_{\{\bar{y}_{T+1} \geq 0\}} \bar{y}_{T+1}$$

- Positive forecast with Gordon-Hall (**PF-GH**)

$$y_{T+1}^{PF-GH} = \frac{1}{B_1} \sum_{j=1}^{B_1} \mathbb{1}_{\{\hat{y}_{T+1}^{*(j)} \geq 0\}} \hat{y}_{T+1}^{*(j)}$$

Assumptions

Assumption

Make the following assumptions on the error process and on the regressor series.

1. *The* ε_t *have mean zero and variance* $\sigma^2 = Var(\varepsilon_t) < \infty$.
2. *The* ε_t *satisfy the Lyapunov-condition* $\mathbb{E}|\varepsilon_t|^{2+\epsilon} \leq C$ *for some* $C, \epsilon > 0$.
3. *The regressor time series has finite mean* $\mathbb{E}x$ *and variance* $Var(x)$.

Applicability of Gordon and Hall (2008) to Forecasting

Proposition

$\tilde{\theta}$ least-squares estimator. Z standard normal random variable with density $\phi(z)$ and cdf $\Phi(z)$. Consider $\theta = \beta$ and $\theta = \mathbb{E}(y_{T+1}|x_T)$ such that $\theta > \theta_1 = 0$. θ_0 population parameter.

1. *Case $\theta_0 > 0$:*
$$\hat{\theta} = \tilde{\theta} + O(T^{-1}).$$
2. *Case $\theta_0 = 0$:*
$$T^{\frac{1}{2}}\tau^{-1}(\hat{\theta} - \theta_0) \xrightarrow{d} Z\Phi(Z) + \phi(Z),$$
τ^2 variance of $\tilde{\theta}$.

Variance Reduction

- Asymptotic distribution of simple constrained estimator $\bar{\theta}$: standard normal truncated to the positive half-line. Variance $(1 - 1/\pi)/2 \approx 0.3408$.
- Distribution of $Z\Phi(Z) + \phi(Z)$ has variance $1/3 + \sqrt{3}/(2\pi) - 1/\pi \approx 0.2907$.
- Thus, in the binding case, $\hat{\theta}$ has about 15% less variance than $\bar{\theta}$.

Variance Reduction for Gordon-Hall Forecast

Proposition

Bagging the positive coefficient forecast $\bar{y}_{t+1}$ (B-PC) is equivalent to computing the forecast $\hat{y}_{t+1}$ from the Gordon-Hall estimator $\hat{\theta}$ (PC-GH).

Shows how and why bagging improves the positivity-restricted forecast $\bar{y}_{t+1}$: variance reduction result from Proposition 1 carries over.

Variance Reduction for Gordon-Hall Forecast 2

Proof.

$$\frac{1}{B}\sum_{j=1}^{B}\bar{y}_{t+1}^{(j)} = \frac{1}{B}\sum_{j=1}^{B}(\bar{\alpha}^{*(j)} + \bar{\beta}^{*(j)}x_t)$$
$$= \left(\frac{1}{B}\sum_{j=1}^{B}\bar{\alpha}^{*(j)}\right) + \left(\frac{1}{B}\sum_{j=1}^{B}\bar{\beta}^{*(j)}\right)x_t$$
$$= \hat{\alpha} + \hat{\beta}x_t = \hat{y}_{t+1}.$$

□

Mean-Squared Error for Gordon-Hall Forecast

Proposition

Let $\tilde{\theta}$ be an unbiased and asymptotically normal estimator of θ with population value $\theta_0 = 0$. Let $\hat{\theta}$ be the Gordon and Hall (2008) bagging estimator of θ. Then,

$$\mathbb{E}(\hat{\theta} - \theta_0)^2 \approx 1.012\mathbb{E}(\tilde{\theta} - \theta_0)^2.$$

A possible advantage of the Gordon and Hall estimator over an unbiased and asymptotically normal estimator in terms of mean-squared error is restricted to finite samples.

Mean-Squared Error for Gordon-Hall Forecast 2

Proof.

$$\mathbb{E}(\hat{\theta}-\theta_0)^2 = \mathsf{Var}(\hat{\theta}) + \mathsf{bias}(\hat{\theta},\, \theta_0)^2.$$

$$\mathsf{bias}(\hat{\theta},\, \theta_0) = \mathbb{E}\hat{\theta} - \theta_0 = \mathbb{E}\hat{\theta}$$

$$\mathbb{E}\hat{\theta} = \frac{1}{B}\sum_{j=1}^{B} \mathbb{E}\mathbb{1}_{\{\tilde{\theta}^{*(j)}\geq 0\}}\mathbb{E}|\tilde{\theta}^{*(j)}| = \frac{1}{2}\mathbb{E}|\tilde{\theta}^{*}| = \frac{1}{2}\sqrt{\frac{2\,\mathsf{Var}(\tilde{\theta}^{*})}{\pi}}.$$

Since $\mathsf{Var}(\tilde{\theta}^{*}) = \mathsf{Var}(\tilde{\theta})$,

$$\begin{aligned}\mathbb{E}(\hat{\theta}-\theta_0)^2 &= \mathsf{Var}(\hat{\theta}) + \frac{\mathsf{Var}(\tilde{\theta})}{2\pi}, \\ &= \frac{\frac{1}{3}+\frac{\sqrt{3}}{2\pi}-\frac{1}{\pi}}{\frac{1}{2}-\frac{1}{2\pi}}\mathsf{Var}(\tilde{\theta}) + \frac{1}{2\pi}\mathsf{Var}(\tilde{\theta}), \\ &\approx 1.012\,\mathsf{Var}(\tilde{\theta}) = 1.012\mathbb{E}(\tilde{\theta}-\theta_0)^2. \qquad \square\end{aligned}$$

A Brief Review of the Literature (1/4)

The Early Days: Evidence of Return Predictability

- Early 1920s: Dow (1920) studied the role of dividend ratios as a possible predictor for returns.
- The 1980s: Fama and Schwert (1977,1981); Rozeff (1984); Keim and Stambaugh (1986); Campbell (1987); Campbell and Shiller (1988a,b); and Fama and French (1988,1989).
  - Key predictors: Lagged values of dividend-price ratio and dividend yield, earnings-price ratio and dividend-earnings ratio, interest rates and spreads, inflation rates, book-to-market ratio, volatility, investment-capital ratio, consumption, wealth, and income ratio, and aggregate net or equity issuing activity.

A Brief Review of the Literature (2/4)

Is Predictability an Illusion?

- Favorable results do not hold during the bull market (1990s): Lettau and Ludvigson (2001) and Schwert (2002).
- Out-of-sample success restricted to some sub-samples: Pesaran and Timmermann (1995).
- Predictability is a statistical illusion: Bossaerts and Hillion (1999).
- Spurious predictability: Nelson and Kim (1993), Cavanagh, Elliott, and Stock (1995), Foster, Smith, and Whaley (1997), Stambaugh (1999), and Ferson, Sarkissian, and Simin (2003).

A Brief Review of the Literature (3/4)

Possible Solutions

- Time-varying models: Pesaran and Timmermann (2002) and Timmermann (2007).
  "Forecasters of stock returns face a moving target that is constantly changing over time..."
- Valid inference: Cavanagh, Elliott, and Stock (1995), Mark (1995), Kilian (1999), Lewellen (2004), Torous, Valkanov, and Yan (2004), Ang and Bekaert (2006), Campbell and Yogo (2006), Jansson and Moreira (2006), Polk, Thompson, and Vuolteenaho (2006), and Elliott (2008).

A Brief Review of the Literature (4/4)

Are the Results Still Inconclusive?

- Goyal and Welch (2008) and Butler, Grullon, and Weston (2006): no out-of-sample predictability.
- Campbell and Thompson (2008): many predictive regressions beat the historical average return once *restrictions* are imposed on the signs of coefficients and return forecasts.
- Rapach, Strauss, and Zhou (2008): *forecast combination* improves out-of-sample equity premium forecasts, handling possible structural instability.

How much R^2? (Campbell and Thompson, 2008)

- Excess return: $r_{t+1} = \mu + x_t + \varepsilon_{t+1}$, $\text{Var}(x)$, $\text{Var}(\varepsilon) < \infty$.
- Investor with single-period horizon, mean-variance preferences

$$U = \text{expected return} - \frac{\gamma}{2} \text{ portfolio variance.}$$

- Without observing x_t: Portfolio weight

$$w = \frac{1}{\gamma} \frac{\mu}{\sigma_x^2 + \sigma_\varepsilon^2}.$$

- Average excess return

$$\frac{1}{\gamma} \frac{\mu^2}{\sigma_x^2 + \sigma_\varepsilon^2} = \frac{S^2}{\gamma},$$

S Sharpe-ratio.

How much R^2? (Campbell and Thompson, 2008)

- Observing x_t: Portfolio weight

$$w_t = \frac{1}{\gamma}\frac{\mu + x_t}{\sigma_\varepsilon^2}.$$

- Average excess return

$$\frac{1}{\gamma}\frac{\mu^2 + \sigma_x^2}{\sigma_\varepsilon^2} = \frac{1}{\gamma}\frac{S^2 + R^2}{1 - R^2}, \; R^2 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2}.$$

- Proportional increase in expected return from observing x_t:

$$\frac{R^2}{1 - R^2}\frac{1 + S^2}{S^2} > \frac{R^2}{S^2}.$$

How much R^2? (Campbell and Thompson, 2008)

- Campbell-Thompson data set (monthly 1871–2005): $S^2 = 0.012$. Out-of-sample R^2 for earnings-price ratio, e.g., **0.0043**.
- Proportional increase in average portfolio return is larger than

$$\frac{R^2}{S^2} = \frac{0.0043}{0.0120} = 0.36 \text{ or } 36\%.$$

Out-of-Sample R^2

We report out-of-sample R^2 statistics multiplied by 100:

$$100R^2_{OS} = 100\left(1 - \frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{T}(y_t - \bar{y}_t)^2}\right),$$

Simulations

$$y_{t+1} = \alpha + \beta x_t + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

- Signal-to-noise β/σ, $\beta = (2^0, 2^{-1}, 2^{-2}, \ldots, 2^{-7})$, $\sigma = 1$.
- Regressor persistence: AR(1) process with different parameters.

$100 \cdot R^2,\ \beta = 1$

AR(1)	0	0.3	0.5	0.8	0.9
UF	3.4074	4.0077	4.6158	10.5224	20.2218
PC	3.4220	4.0193	4.6252	10.5226	20.2229
PC-GH	3.4634	4.0547	4.6515	10.5238	20.2232
PF	3.4074	4.0085	4.6166	10.5225	20.2237
PF-GH	3.3980	4.0048	4.6134	10.5157	20.2250

Simulations 2

$100 \cdot R^2, \ \beta = 1/2$

AR(1)	0	0.3	0.5	0.8	0.9
UF	0.2147	0.4330	0.5870	2.3659	5.4493
PC	0.3323	0.5602	0.6848	2.4097	5.4683
PC-GH	0.4065	0.6525	0.7726	2.4859	5.5112
PF	0.2252	0.4418	0.5989	2.3823	5.4880
PF-GH	0.2476	0.4654	0.6177	2.4046	5.5163

Simulations 3

$100 \cdot R^2, \; \beta = 1/4$

AR(1)	0	0.3	0.5	0.8	0.9
UF	-0.4734	-0.4861	-0.2177	-0.1571	0.7566
PC	-0.1988	-0.1922	0.0313	0.0154	0.9247
PC-GH	-0.1659	-0.1364	0.0689	0.0862	1.0222
PF	-0.3607	-0.3788	-0.1339	-0.0493	0.8688
PF-GH	-0.2503	-0.2610	-0.0435	0.0559	0.9936

Simulations 4

$100 \cdot R^2, \ \beta = 1/8$

AR(1)	0	0.3	0.5	0.8	0.9
UF	-0.6075	-0.6761	-0.6788	-0.6255	-0.5323
PC	-0.2492	-0.3086	-0.3221	-0.2519	-0.1885
PC-GH	-0.2545	-0.3307	-0.3360	-0.2321	-0.1538
PF	-0.3702	-0.3638	-0.3591	-0.2576	-0.2138
PF-GH	-0.2269	-0.2219	-0.2227	-0.1162	-0.0452

Data

- Monthly data from Campbell and Thompson (2008).
- Excess returns on the S&P 500 (1871M2 - 2005M12) over the 3-month Treasury-Bill interest rate (1870M1 - 2005M12).
- The predictor variables are the dividend yield (**d/p**, 1872M2 - 2005M12), earnings yield (**e/p**, 1872M2 - 2005M12), smoothed earnings yield (**se/p**, 1881M1 - 2005M12), book-to-market ratio (**b/m**, 1926M6 - 2005M12), smoothed return on equity (**roe**, 1936M6 - 2005M12), the 3-month Treasury-Bill (**tbl**, 1920M1 - 2005M12), long-term government bond yield (**lty**, 1870M1 - 2005M12), the term spread (**ts**, 1920M1 - 2005M12), the default spread (**ds**, 1919M1 - 2005M12), the lagged inflation rate (**inf**, 1871M5 - 2005M12), the equity share of new issues (**nei**, 1927M1 - 2005M12), and the consumption-wealth ratio (**cay**, 1951M12 - 2005M12).

Sign Restrictions

Variable	Sign β	Variable	Sign β
d/p	$+$	**lty**	-
e/p	$+$	**ts**	$+$
se/p	$+$	**ds**	$+$
b/m	$+$	**inf**	-
roe	$+$	**nei**	-
tbl	-	**cay**	$+$,-,-

Results 1

	d/p	e/p	se/p	b/m	roe	tbl
Begin						
Sample	1872:1	1872:2	1881:2	1926:6	1936:6	1920:1
Forecast	1927:1	1927:1	1927:1	1946:1	1956:6	1940:1
Campbell and Thompson (2008)						
IS R^2	1.1208	0.7081	1.3521	0.6078	0.0225	0.8691
IS t	1.2519	2.2829	1.8495	1.9566	0.3587	2.4586
UF	-0.6563	0.1159	0.3239	-0.3968	-0.9259	0.5048
PC	0.0483	0.1770	0.4175	-0.3968	-0.0778	0.4899
PF	0.0723	0.1321	0.3777	0.0168	-0.9259	0.5401
PCF	0.0798	0.1770	0.4297	0.0168	-0.0778	0.5251
i.i.d. bootstrap						
B-PC	0.2644	0.2908	0.4319	-0.5697	-0.5852	0.5723
B-PF	0.5060	0.2433	0.4529	0.2728	-0.7925	0.5468
B-PCF	0.3558	0.2992	0.4894	0.0943	-0.2446	0.3640
moving block bootstrap						
B-PC	0.1277	0.1703	0.5128	-1.1542	-0.7427	0.4207
B-PF	0.2649	0.1696	0.5494	0.2259	-1.0368	0.3923
B-PCF	0.2677	0.2174	0.5853	-0.2123	-0.3855	0.2372
Gordon and Hall (2008)						
PC-GH iid	0.7388	0.3829	0.6704	-0.7287	-1.1179	0.5999
PC-GH mbb	0.3172	0.2655	0.5433	-2.0418	-0.7579	0.5547
PF-GH iid	0.8865	0.3850	0.6884	-0.0454	-1.1179	0.6046
PF-GH mbb	0.4134	0.2902	0.5649	-0.1343	-0.7579	0.5809

Results 2

	lty	*ts*	*ds*	*inf*	*nei*	*cay*
Begin						
Sample	1870:1	1920:1	1919:1	1871:5	1927:12	1951:12
Forecast	1927:1	1940:1	1939:1	1927:1	1947:12	1971:12
Campbell and Thompson (2008)						
IS R^2	0.1911	0.6561	0.0956	0.0549	0.4828	2.5983
IS t	1.4730	2.1804	0.7248	0.3898	1.7510	3.5924
UF	-0.1893	0.4449	-0.2391	-0.2210	0.3218	-1.5393
PC	-0.1893	0.4565	-0.2391	-0.2092	0.3218	-1.5393
PF	0.2047	0.4392	-0.2391	-0.1834	0.4843	0.1596
PCF	0.2047	0.4507	-0.2391	-0.1717	0.4843	0.1596
i.i.d. bootstrap						
B-PC	-0.1783	0.5558	-0.5579	-0.1296	0.3388	-1.2483
B-PF	0.2178	0.2548	0.0802	-0.0329	0.3921	0.4458
B-PCF	0.1676	0.2530	-0.4871	-0.1430	0.4070	0.4490
moving block bootstrap						
B-PC	-0.2970	0.3320	-0.4614	-0.3859	0.1342	-2.9017
B-PF	0.1788	0.0665	0.0202	-0.2413	0.2762	-0.7165
B-PCF	0.1495	0.0266	-0.3868	-0.3567	0.3146	-0.5658
Gordon and Hall (2008)						
PC-GH iid	-0.1544	0.7707	-0.9398	-0.0728	0.3276	-1.3136
PC-GH mbb	-0.2229	0.6153	-0.6771	-0.2314	0.2210	-2.9282
PF-GH iid	0.2520	0.7130	-0.6272	0.0146	0.4901	0.2914
PF-GH mbb	0.2388	0.5902	-0.4807	-0.0368	0.3998	-0.9646

Conclusions

- Consider imposing restrictions in a univariate regression forecast model.
- Propose bagging estimator of Gordon and Hall for forecast situation and show its applicability.
- In simulations, find that small-sample gains are obtained for a restricted region of signal-to-noise ratio and regressor persistence.
- Applying the procedure to Campbell and Thompson (2008) data set, we find forecast improvements.