SDU❧

Two out-of-sample forecasting models of the equity premium

by

Thiago de Oliveira Souza

Two out-of-sample forecasting models of the equity premium

Thiago de Oliveira Souza

University of Southern Denmark;

Danish Finance Institute

October 27, 2020

Abstract

I derive two valid forecasting models of the equity premium in monthly frequency, based on little more than no-arbitrage: A "predictability timing" version of partial least squares, given that predictability is theoretically time-varying; and a least squares model with realized market premiums in monthly frequency as the regressor, since realized returns are theoretically correlated to risk and to the price of risk. This evidence is consistent with the instability inherent to monthly equity premium forecasts based on standard partial least squares and disaggregated book-to-markets as regressors, and with the fact that taking one extra lag of book-to-markets in predictive return regressions improves the estimates.

1 Introduction

Partial least squares (PLS) based on disaggregated book-to-market (BM) ratios (Kelly and Pruitt, 2013) cannot generate a valid forecasting model of the monthly equity premium according to the hurdle proposed by Welch and Goyal (2008), such as positive out-of-sample (OOS) performance since the Oil Shock (de Oliveira Souza, 2020b). The contribution of the present paper is to discuss why this happens in terms of standard asset pricing theory and especially to document two valid models, instead.

Model 1 simply predicts the monthly market premium, MP_t, from its (one-month) lag,

$$\mathrm{E}_t\left[MP_{t+1}\right] = \alpha + \beta \, MP_t, \tag{1}$$

where α and β are estimated by ordinary least squares. This supports the explanation that PLS return forecasts based on BMs with one extra lag are more accurate because the market premium is unconditionally positively autocorrelated for a one-month lag (de Oliveira Souza, 2020b).

Model 2 implements a "predictability timing" version of the original PLS procedure: The recursive prediction of the (future) market premium in time t is the historical mean, $\overline{MP}_t$, if the absolute forecasting error realized in time t for the historical mean, $|\overline{MP}_{t-1} - MP_t|$, is smaller than the one for the PLS model, $|\widehat{MP}_{pls,t-1} - MP_t|$. Otherwise, I use the PLS prediction, $\widehat{MP}_{pls,t}$,

$$|\overline{MP}_{t-1} - MP_t| < |\widehat{MP}_{pls,t-1} - MP_t| \implies \mathrm{E}_t\left[MP_{t+1}\right] = \overline{MP}_t, \tag{2}$$

$$|\overline{MP}_{t-1} - MP_t| \geq |\widehat{MP}_{pls,t-1} - MP_t| \implies \mathrm{E}_t\left[MP_{t+1}\right] = \widehat{MP}_{pls,t}. \tag{3}$$

The paper proceeds as follows: Section 2 presents the data and the OOS performance measures that I use to evaluate the models. Sections 3 and 4, respectively, for models 1 and 2, present the theoretical and empirical frameworks in detail, the in-sample and OOS

evidence, in addition to the economic performance of simple investment strategies based on the models. Section 5 concludes.

2 Data and forecasting performance

BM data is for the 6 portfolios double sorted by market equity (ME) and BM in Kenneth French library. The steps in de Oliveira Souza (2020b) explain how to obtain annual book equity (BE) from June each year and market equity each month. The BMs are in logs.

The market premium is the monthly return on the Center for Research in Security Prices (CRSP) index portfolio minus the risk-free rate from Kenneth French library, which I transform to continuous compounding. The period is June 1926 to December 2019.

2.1 OOS evaluation

Most of the evidence that I present involves running all regressions recursively on training samples that exclude the return that is ultimately forecast for a large number of split dates. I follow Welch and Goyal (2008) and Kelly and Pruitt (2013), for example, and also report the results graphically by split date.

The only quantitative test: Inference about predictability is based on the forecast encompassing test for nested models of Clark and McCracken (2001), ENC-NEW. I report the test statistics – along with their respective critical values – over a wide range of split dates to analyze how robust the conclusions are. The null hypothesis in the test is that the extra regressors in the larger models (relative to the sample mean) have no predictive content. A rejection validates the regressors.

OOS R^2**:** The ENC-NEW test is not exactly based on the mean squared error (MSE) of the forecasts. For this qualitative assessment, I report the OOS R^2, given by

$$OOS\,R^2 = 1 - \frac{\sum_t (y_t - \widehat{y}_t)^2}{\sum_t (y_t - \overline{y}_t)^2},\tag{4}$$

where the model's prediction is $\widehat{y}_t$, the historical mean is $\overline{y}_t$, and the realization is y_t. In finite samples, the OOS R^2 is negative under the null hypothesis of no predictability, **not** zero (Clark and West, 2006). Hence, a negative OOS R^2 is not the same as no predictability: The OOS R^2 is a measure of forecasting accuracy, but not a predictability test.

Cumulative ΔSSE**:** Another, more flexible, qualitative performance measure is the cumulative sum of squared forecasting errors of the historical mean minus the one from each model, ΔSSE. I report realized values at each date, as in Welch and Goyal (2008). Intuitively, increases in ΔSSE between two dates imply that the model in question predicts better than the historical mean *for that time interval*. Hence, unlike the OOS R^2, it is possible to freely adjust starting and ending dates, and to check the performance of the models in different periods.

3   Model 1: OLS forecast based on realized premiums

This section starts by presenting a general framework for analyzing time-varying risk premiums (de Oliveira Souza, 2020a; Munk, 2013). The main theoretical restriction is the absence of arbitrage, so that a stochastic discount factor (SDF) exists.

Let $P_i = (P_{i,t})$ be the price of asset $i = 1, ..., I$, following the process

$$dP_{i,t} = P_{i,t}\left[\mu_{i,t}\,dt + \boldsymbol{\sigma}_{i,t}^{\top}d\boldsymbol{z}_t\right],\tag{5}$$

where $z = (z_t)$ is a standard Brownian motion of dimension d, representing different shocks; $\sigma_{i,t}$ is a vector of sensitivities to these exogenous shocks; and $\top$ is the transposition sign (boldfaces denote vectors).

Definition 1 *In a continuous-time setting in which the fundamental uncertainty is represented by $z = (z_t)$ with dimension d and a locally risk-free asset with return r_t^f is traded at all times, **the price of risk** is a process $\lambda = (\lambda_t)$ with dimension d, satisfying*

$$\mu_{i,t} - r_t^f = \sigma_{i,t}^\top \lambda_t \tag{6}$$

*for all assets $i = 1, ..., I$. Alternatively, in a continuous-time setting in which the fundamental uncertainty is represented by a one-dimensional Brownian motion, $z = (z_t)$, the one-dimensional (quantity of) **risk** is a time-varying scalar given by*

$$\sigma_{i,t}^{mv} = \frac{\sigma_i^\top \lambda_t}{\|\lambda_t\|}, \tag{7}$$

where $\|.\|$ is the Euclidean norm and $\|\lambda_t\|$ is the instantaneous Sharpe ratio of the mean-variance frontier (MVF).

Indeed, this also defines the SDF (as formally proven by Munk, 2013, pp. 109-112):

Proposition 1 *Under technical conditions and in a continuous-time setting in which the fundamental uncertainty is represented by $z = (z_t)$ with dimension d, a process $\zeta = (\zeta_t)$ is a SDF if and only if*

$$d\zeta_t = -\zeta_t \left[r_t^f \, dt + \lambda_t^\top dz_t \right], \quad \zeta_0 = 1. \tag{8}$$

Equivalently, if the fundamental uncertainty is represented by a one-dimensional Brownian motion, $z = (z_t)$, we obtain

$$d\zeta_t = -\zeta_t \left[r_t^f \, dt + \|\lambda_t\| dz_t \right], \quad \zeta_0 = 1. \tag{9}$$

Intuitively, the (d-dimensional) price of risk in Eq. (6), $\boldsymbol{\lambda}_t$, is the compensation in terms of excess expected return per unit of risk stemming from each of d economic shocks. The d-dimensional volatility parameter in the same equation, $\boldsymbol{\sigma}_{i,t}$, determines how sensitive the price of each individual asset is to such shocks. However, $\boldsymbol{\sigma}_{i,t}$ only determines the composition of the shocks to which the asset is exposed; not its "quantity" of risk. In contrast, the one-dimensional risk, $\sigma_{i,t}^{mv}$ in Eq. (7), gives this information by aggregating the risks stemming from all shocks, conditioned on their prevailing risk prices, into a single value. Hence, $\sigma_{i,t}^{mv}$ is the (quantity of) risk of the asset in a more intuitive sense.

Lemma 1 *Under the assumption that the conditional risk of the market portfolio is independent of the price of risk, realized market returns are positively correlated with the conditional market risk and with the price of risk.*

Proof: *Let $\sigma_{mp,t}^{mv}$ be the one-dimensional conditional risk of the market portfolio. The realized excess return on the market in discrete time, based on Eqs. (6) and (7), is then*

$$MP_t \equiv \mu_{mp,t-1} - r_{t-1}^f + \epsilon_t = \sigma_{mp,t-1}^{mv}\|\boldsymbol{\lambda}_{t-1}\| + \epsilon_t, \tag{10}$$

where ϵ_t is an error term. If the risk of the market portfolio is independent of the price of risk, the correlations between the realized equity premium and the (lagged) market risk or price of risk become, respectively,

$$\rho\left(MP_t, \sigma^{mv}_{mp,t-1}\right) = \frac{cov\left(\sigma^{mv}_{mp,t-1}\|\boldsymbol{\lambda}_{t-1}\| + \epsilon_t, \sigma^{mv}_{mp,t-1}\right)}{\sqrt{var\left(MP_t\right) var\left(\sigma^{mv}_{mp,t-1}\right)}}$$

$$= \frac{\mathrm{E}[\|\boldsymbol{\lambda}_{t-1}\|]\sqrt{var\left(\sigma^{mv}_{mp,t-1}\right)}}{\sqrt{var\left(MP_t\right)}} > 0, \tag{11}$$

$$\rho\left(MP_t, \|\boldsymbol{\lambda}_{t-1}\|\right) = \frac{cov\left(\sigma^{mv}_{mp,t-1}\|\boldsymbol{\lambda}_{t-1}\| + \epsilon_t, \|\boldsymbol{\lambda}_{t-1}\|\right)}{\sqrt{var\left(MP_t\right) var\left(\|\boldsymbol{\lambda}_{t-1}\|\right)}}$$

$$= \frac{\mathrm{E}\left[\sigma^{mv}_{mp,t-1}\right]\sqrt{var\left(\|\boldsymbol{\lambda}_{t-1}\|\right)}}{\sqrt{var\left(MP_t\right)}} > 0, \tag{12}$$

given that the market portfolio is risky and the price of risk is theoretically positive. ∎

3.1 Empirical evidence in sample

Without further assumptions, Lemma 1 implies that realized market returns are also correlated with future market returns for as long as market risk and (or) the price of risk are persistent (positively autocorrelated). Indeed, several theoretically micro-founded asset pricing theories generate low frequency changes in both variables. Notably, these changes happen at the business-cycle frequency in consumption-based models (as in Cochrane, 2017). Indeed, Table 1 suggests that the variables are persistent at the monthly frequency.

Table 1 estimates regressions similar to Eq. (1) for $h = 1$ or $h = 2$ months ahead,

$$MP_{t+h} = \alpha + \beta MP'_t + \epsilon_{t+h}, \tag{13}$$

Table 1: Autoregressive forecasts of the monthly market premium (MP, in %) one or two periods ahead from June 1926 to December 2019.

	(1) MP_{t+1}	(2) MP_{t+1}^{low}	(3) MP_{t+1}^{high}	(4) MP_{t+1}	(5) MP_{t+2}
MP_t	0.10	0.06	0.18		-0.01
	(1.94)	(0.69)	(1.77)		(-0.24)
MP_t^+				0.18*	
				(2.17)	
Constant	0.46**	0.34	0.14	0.11	0.51**
	(2.83)	(0.93)	(0.33)	(0.49)	(2.97)
R^2	0.010	0.002	0.016	0.011	0.000
t-ols(β)	3.41	0.86	3.32	3.60	-0.36
Months	1121	433	688	1121	1120

I forecast the monthly equity premium $h = 1$ or $h = 2$ months ahead, with equations of the form

$$MP_{t+h} = \alpha + \beta \, MP_t' + \epsilon_{t+h}$$

where MP_t' is the realized market premium in all but model (4), which sets negative realizations of the market premium to zero. In (2)/(3) the estimates are conditional on negative/positive market premiums (as a regressor). Coefficients with Newey and West (1987) standard errors with 1 lag, except models (2) and (3) that use automatic lags (Newey and West, 1994)), t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The lower panel shows the uncorrected t statistics, t-ols(β), the number of months, and the R^2.
Summary: Realized excess returns – especially when positive – are positively correlated with one period ahead excess returns at the monthly frequency.

where MP_t' is simply the market premium in every model, except in (4):[1] The marginally

significant slope in (1) confirms that present and future market premiums are (marginally)

positively correlated in the short run. In model (2) (model (3)), I estimate Eq. (13)

exclusively when MP_t' is negative (positive): Negative realized market premiums are

largely uncorrelated with future premiums, in (2); while positive values are more correlated,

although insignificant, in (3).[2] Indeed, this is expected because risk and price of risk –

which realized returns proxy for in this equation, according to Lemma 1 – are theoretically

---

[1]All models use Newey and West (1987) standard errors with 1 lag, but models (2) and (3) use automatic lags (Newey and West, 1994).

[2]It is difficult to control for autocorrelation in (2) and (3), because both series have missing values. Indeed, the uncorrected t-statistic is highly significant for model (3).

non-negative. Confirming the theory, model (4) sets negative realized premiums to zero and improves the forecasts.

Finally, (5) is equivalent to (1) for $h = 2$: The present equity premium and the premium two months ahead are uncorrelated. This supports the explanation that BMs with one extra lag are better return predictors at the monthly frequency because the extra lag removes the positive autocorrelation in returns that affects the estimation (de Oliveira Souza, 2020b).

3.2 Empirical evidence out of sample
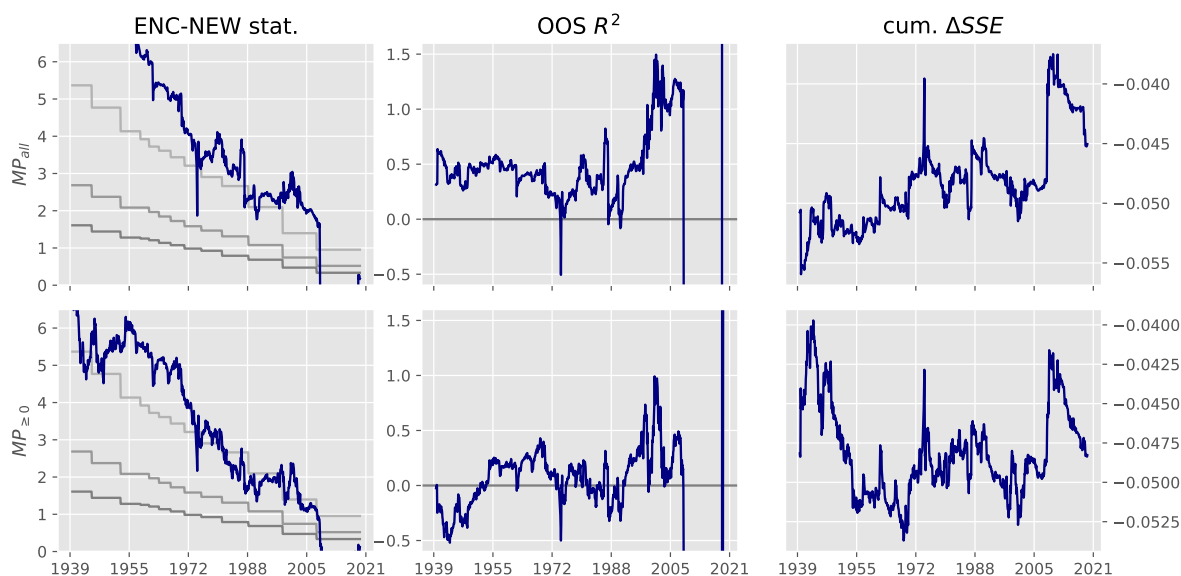
Figure 1: Out-of-sample statistics by sample split date for recursive OLS forecasts of monthly market premiums based on their lagged realizations from June 1926 to December 2019.

The graphs show (left) ENC-NEW test statistics of Clark and McCracken (2001), with gray kinked lines for the critical values at 1, 5, and 10 percent (from light to dark); (center) OOS R^2 by sample split date; and (right) the cumulative sum of squared forecasting errors of the historical mean minus the one from each model, ΔSSE. There are more extreme values outside the bounds in each graph (unreported for scaling reasons). The top model, MP_{all}, uses all realizations of the market premium as a regressor; the bottom model, $MP_{\geq 0}$, sets negative realizations to zero and use the result as a regressor instead.
Summary: Both models have significant (and relatively similar) ENC-NEW statistics. But model MP_{all} has larger OOS R^2 (positive for almost any date before 2010), while ΔSSE increases more or less monotonically since 1940: It largely overcomes the hurdle of Welch and Goyal (2008), even after 2000 (until 2010).

Fig. 1 confirms Lemma 1 and the performance of models (1) and (4) out of sample. The standard autoregressive model, MP_{all} in Fig. 1 (column (1) in Table 1), largely overcomes the hurdle of Welch and Goyal (2008) even after 2000 (until 2010): For almost any split date between 1940 and 2010, the monthly forecasts deliver positive OOS R^2, the ENC-NEW statistic is significant at the 1 percent level, and ΔSSE increases more or less monotonically, exactly as required. The only drawback is a large spike in predictability around 1975 (clear in the ΔSSE graph), which is associated with subsequent under-performance. Hence, the OOS R^2 is negative for this split date. In addition, the model with negative premium realizations set to zero, $MP_{\geq 0}$ in Fig. 1 (column (4) in Table 1), has slightly less significant ENC-NEW statistics for the same date interval. Hence, both are valid forecasting models. However, there are relatively large differences in terms of qualitative measures: The OOS R^2 is negative for several split dates, and the upward drift in ΔSSE is almost non-existent.

3.3 Economic performance out of sample

The evidence above is also economically significant. In this section, I partition the sample into months following positive, MP^{high}, or negative, MP^{low}, market premium realizations, and evaluate the premiums in these two groups separately. Table 2 reports descriptive statistics for this analysis in the full sample or for samples that start in 1980 ($_{80+}$), 2000 ($_{00+}$), or 2010 ($_{10+}$).

The main results in Table 2 are in columns (1) and (5), for the full sample. The monthly equity premium is small and insignificant after negative market premium realizations, 0.1% (0.32 t statistic) with 0.05 Sharpe ratio, in (1). The premium is more than seven times larger and strongly significant after positive realizations, 0.78% (4.32) with an impressive 0.57 Sharpe ratio, in (5). Skewness is similar, but kurtosis is higher in the high premium periods, in (5). Finally, returns are positive about 60% of the time in the full sample (689 out of 1122 months correspond to MP^{high}). The annualized Sharpe ratio is the monthly Sharpe ratio multiplied by $\sqrt{12}$.

The evidence in more recent periods – (2) to (4) for MP^{low}, and (6) to (8) for MP^{high} – largely mirror the one for the full sample. The only difference appears very recently, since 2010, with the important caveat that these are relatively small samples: 30 months in (4), and 78 months in (8). In this period, the two groups of premiums have essentially the same mean, but the MP^{high} Sharpe ratio of 1.15 in (8) is twice the one of MP^{low} in (4), 0.57. In fact, there are also large differences in skewness and kurtosis in this smaller sample.

Table 2: Market premium following positive or negative market premium realizations in July 1926 – December 2019 (and subperiods).

	(1) MP^{low}	(2) 80+	(3) 00+	(4) 10+	(5) MP^{high}	(6) 80+	(7) 00+	(8) 10+
μ_y	0.10	0.19	0.01	0.80	0.78***	0.73***	0.68*	0.89**
	(0.32)	(0.45)	(0.02)	(0.90)	(4.32)	(3.47)	(2.28)	(2.92)
SR_y	0.05	0.12	0.01	0.57	0.57	0.70	0.65	1.15
γ_1	-0.46	-1.14	-0.74	-0.18	-0.51	-0.49	-0.89	-1.13
α_4	5.71	6.26	4.30	2.47	15.97	3.70	4.48	5.50
Months	433	175	80	30	689	293	148	78

Columns (1) to (4) report statistics for the months that follow negative market premiums; months in (5) to (8) follow positive premiums. I repeat the evaluation of (1)/(5) in subsamples after 1980, 2000, and 2010 (shown as $_{80+}$, $_{00+}$, and $_{10+}$). μ_y is the annualized mean return, standard OLS t statistics in parentheses, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. SR_y is the monthly Sharpe ratio multiplied by $\sqrt{12}$, γ_1 is skewness, α_4 is the kurtosis.
Summary: Returns and Sharpe ratios are substantially larger in months that follow positive market premium realizations, being close to zero otherwise.

4 Model 2: PLS predictability timing

One of the main theoretical reasons why PLS forecasts fail to provide the unconditional evidence required by Welch and Goyal (2008) is the fact that predictability theoretically concentrates in bad times within a standard risk-based asset pricing framework (de Oliveira Souza, 2020c).[3] Hence, the performance of every forecasting model must be particularly positive

---

[3]Cujean and Hasler (2017) reach the same conclusion, but require individuals to make systematic mistakes to remain valid.

in bad times, but not otherwise: The irregular PLS performance out of sample – which fails the stability tests in Welch and Goyal (2008) – does not necessarily disqualify the model.

The intuition in de Oliveira Souza (2020c) is essentially that within *any* risk-based asset pricing framework, realized returns have two components: Predictable risk premiums and unpredictable shocks, as in Eq. (10). In bad times, the price of risk, $\|\boldsymbol{\lambda}_t\|$, increases. Hence, the predictable fraction of returns – and predictability – increases. In addition, the predictions are based on proxy variables for risk or for the price of risk. For price of risk proxies in particular, de Oliveira Souza (2020c) further proves that a second similar effect reinforces the one just described: The fraction of the proxy determined by the latent price of risk also increases relative to the one determined by noise ("measurement error") when the price of risk increases: Changes in the proxy become more informative about changes in the price of risk when the variation in the price of risk is larger if we keep measurement error constant. In summary, the theory in de Oliveira Souza (2020c) suggests that "predictability timing" should improve the performance of the forecasts. The present section implements the specific predictability timing procedure described by Eqs. (2) and (3) in the introduction, and confirms this theoretical prediction.

The underlying PLS prediction: The standard PLS procedure, described in detail by Kelly and Pruitt (2013) and complemented by de Oliveira Souza (2020b), is the following for the market premium:

(i) Run a time-series regression of the **adjusted** BMs of each asset i, $bm'_{i,t}$, on the (future) excess return on the market,

$$bm'_{i,t} = \widehat{\Phi}_{i,0} + \widehat{\Phi}_i MP_{t+h} + e_{i,t}. \tag{14}$$

(ii) Run a cross-sectional regression, for each period, t, of all adjusted BMs (at once) on the loadings estimated in the first stage, $\widehat{\Phi}_i$, in Eq. (14),

$$bm'_{i,t} = \widehat{c}_t + \widehat{F}_t \widehat{\Phi}_i + \omega_{i,t}, \tag{15}$$

which results in an estimate for the latent factor, F_t, for each period.

(iii) Run a standard predictive regression of returns on the factor estimated in the second stage, $\widehat{F}_t$,

$$MP_{t+h} = \beta_0 + \beta \widehat{F}_t + u_{t+h}. \tag{16}$$

The fitted $\widehat{MP}_{t+h}$ from Eq. (16) is the underlying PLS forecast used in Eq. (3), which I estimate recursively.

4.1 BM frequency and preliminary PLS results

In order to run the regressions above, Kelly and Pruitt (2013) pre adjust the series of BMs of each firm i by dividing the BMs by their respective time-series standard deviation, $\sigma_{bm_l,i}$, and possibly include one extra lag (de Oliveira Souza, 2020b),

$$bm'_{m,i,t} \equiv \frac{bm_{m,i,t-1}}{\sigma_{bm_l,i}}, \tag{17}$$

where $\sigma_{bm_l,i}$ is the standard deviation of the BM series already with the extra lag. In the recursive estimation, $\sigma_{bm_l,i}$ becomes time-varying (due to being calculated over the sample that is inside the information set when the forecast is created). These are the BMs that I use for the baseline PLS model.

However, as de Oliveira Souza (2020b) notices, book equity (BE) is fixed from June to June. Hence, BM innovations in all other months are returns on the respective BM/ME portfolios. The pervasive correlation between past and future returns at the monthly frequency that I document in Section 3 implies that past returns have a strong effect on the forecasts. The extra BM lag in Eq. (17) helps to alleviate this effect.

An alternative that completely avoids confounding returns and BMs is to only update the BMs when BE is also updated. This only happens in June. Hence, the adjusted BM is

the BM from the latest June (kept constant for a year), without extra lags, divided by its standard deviation,

$$bm'_{Jun,i,t} \equiv \frac{bm_{Jun,i,t}}{\sigma_{bm_{Jun},i}}, \tag{18}$$

where $\sigma_{bm_{Jun},i}$ is the standard deviation of the BM series with the June values (but $\sigma_{bm_{Jun},i}$ is calculated each month in the recursive implementation and varies with the sample).

Table 3 compares the performance of the standard PLS forecasts (without predictability timing) based on how often the monthly BMs are updated. Table 3 suggests that monthly BMs updated yearly generate slightly better performance: The factor slope in Eq. (16) is marginally significant (0.06 level), the OOS R^2 is positive and the ENC-NEW test is significant for December 1989 as the split date. In contrast, the results are insignificant at the usual level for the PLS forecasts based on BMs updated monthly.

Table 3: Summary of PLS predictions in July 1926 – December 2019.

	IS $R^2(\%)$	$p(NW)$	OOS $R^2(\%)$	$p(ENC)$
BM_m	0.28	0.14	-0.10	<0.10
BM_{Jun}	0.43	0.06	0.36	<0.05

BM_m (BM_{Jun}) corresponds to PLS forecasts based on BMs in monthly (yearly) frequency. In yearly frequency, the BMs are from the latest June. The table reports the IS R^2; p-value of $\widehat{F_t}$ in the last-stage PLS regression with Newey and West (1987) standard errors with 1 lag, $p(NW)$; OOS R^2; and p-value of the ENC-NEW test (Clark and McCracken, 2001), $p(ENC)$. The OOS split date is December 1989.
Summary: The PLS forecasts are significant and slightly more accurate, with positive OOS R^2, when only June BMs are used in the estimation.

4.2 Empirical evidence out of sample

First, the left-hand column in Fig. 2 shows the ENC-NEW test statistics of Clark and McCracken (2001) for the standard PLS forecasts (without predictability timing), with gray kinked lines for the critical values at 1, 5, and 10 percent (from light to dark). These graphs confirm that standard PLS forecasts based on BMs updated only in June, $BM_y(June)$ in

the top row, generate significant OOS performance for a wide range of split dates. The evidence is marginal when BMs are updated monthly, BM_m.

The other two columns in Fig. 2 compare the performance of the standard PLS model ((light) orange) with its respective predictability timing (dark blue) in terms of OOS R^2 and cumulative ΔSSE. The PLS regressors are the BMs of the 6 BM/ME portfolios, updated only in June (Eq. (18), top graphs) or each month (Eq. (17), bottom graphs). These two columns in Fig. 2 confirm that predictability timing (in blue) substantially improves the PLS forecasts regardless of how often BMs are updated. But the performance based on BMs updated in June are particularly positive for almost any split date. In fact, the OOS R^2 for the predictability timing forecasts in both cases is positive since the Oil Shock. The positive drifts in both ΔSSE graphs are also much clearer for the predictability timing forecasts. This finally overcomes the hurdle proposed by Welch and Goyal (2008) to validate both models.

4.3 Economic performance out of sample

I consider three investment scenarios based on predictability timing PLS forecasts:[4] (i) The market premium forecast is above average, MP^{high}; (ii) below average, MP^{low}; or (iii) exactly average, $MP^{neutral}$. Naturally, scenarios (i) and (ii), reported in Table 4a, only happen if the PLS prediction is being used (which means that the PLS delivered small prediction error relative to the historical mean in the preceding month) and scenario (iii), in Table 4b, only happens when the PLS forecast is ignored (for being previously inaccurate). The tables report the market premium mean, t statistics, annualized Sharpe ratio (monthly value multiplied by $\sqrt{12}$), SR_y, skewness, γ_1, and kurtosis, α_4.

The main results are in Table 4a: Columns (1) to (4) report statistics for the low forecast months, and (5) to (8) for high forecasts. The full sample results, in (1) and (5), show that predictability timing PLS forecasts are remarkably useful: The mean market premium in high forecast months is 0.81% (3.93 t statistics) compared to -0.41% (-0.91) in low

---

[4]Appendix A has summary statistics based on standard PLS forecasts, instead.

Table 4: Investment performance of PLS predictability timing forecasts in July 1926 – December 2019 (or recently)

(a) Above, MP^{high}, or below average premium forecasts, MP^{low}

	(1) MP^{low}	(2) 80+	(3) 00+	(4) 10+	(5) MP^{high}	(6) 80+	(7) 00+	(8) 10+
$Timing_\mu$	-0.41	0.20	0.07	1.21	0.81***	0.57	0.59	0.86
	(-0.91)	(0.36)	(0.10)	(1.58)	(3.93)	(1.59)	(1.11)	(0.56)
SR_y	-0.22	0.12	0.04	0.91	0.76	0.57	0.65	1.12
γ_1	-0.84	-1.42	-0.81	-0.35	-0.43	-0.24	-0.67	0.34
α_4	5.16	6.15	4.02	2.78	3.64	3.42	3.38	1.50
Months	203	117	65	36	318	95	35	3

Columns (1) to (4) are low forecast months; (5) to (8) are high. Full sample in (1) and (5), and respective subsamples after 1980, 2000, and 2010 as $_{80+}$, $_{00+}$, and $_{10+}$.

(b) All months, MP^{hist}, or average premium forecasts, $MP^{Neutral}$.

	(1) MP^{hist}	(2) 80+	(3) 00+	(4) 10+	(5) $MP^{neutral}$	(6) 80+	(7) 00+	(8) 10+
$Timing_\mu$	0.51**	0.54**	0.43	0.86*	0.67**	0.68**	0.58	0.67*
	(3.22)	(2.64)	(1.51)	(2.58)	(2.90)	(2.86)	(1.72)	(2.05)
SR_y	0.33	0.42	0.35	0.86	0.41	0.62	0.53	0.86
γ_1	-0.54	-1.08	-0.91	-0.55	-0.27	-0.25	-0.74	-1.28
α_4	10.01	6.97	5.13	4.20	11.39	3.61	4.32	5.31
Months	1122	467	227	107	601	255	127	68

Columns (1) to (4) are all months; (5) to (8) are average forecast months. Full sample in (1) and (5), subsamples after 1980, 2000, and 2010 as $_{80+}$, $_{00+}$, and $_{10+}$.

$Timing_\mu$ is the mean return, OLS t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, SR_y is the annualized Sharpe ratio, γ_1 is skewness, α_4 is the kurtosis. PLS forecasts are based on yearly BMs with no extra lags.

Summary: Full sample returns and Sharpe ratios are substantially larger in high forecast months than in low ones (Table 4a), and they are average (between these extremes) in average forecast months (Table 4b). High forecasts are the majority in the full sample (318/203) but decrease over time (Table 4a). In addition, relatively inaccurate PLS forecasts increase in proportion over time from 601/1122 in the full sample to 68/107 since 2010 (Table 4b).
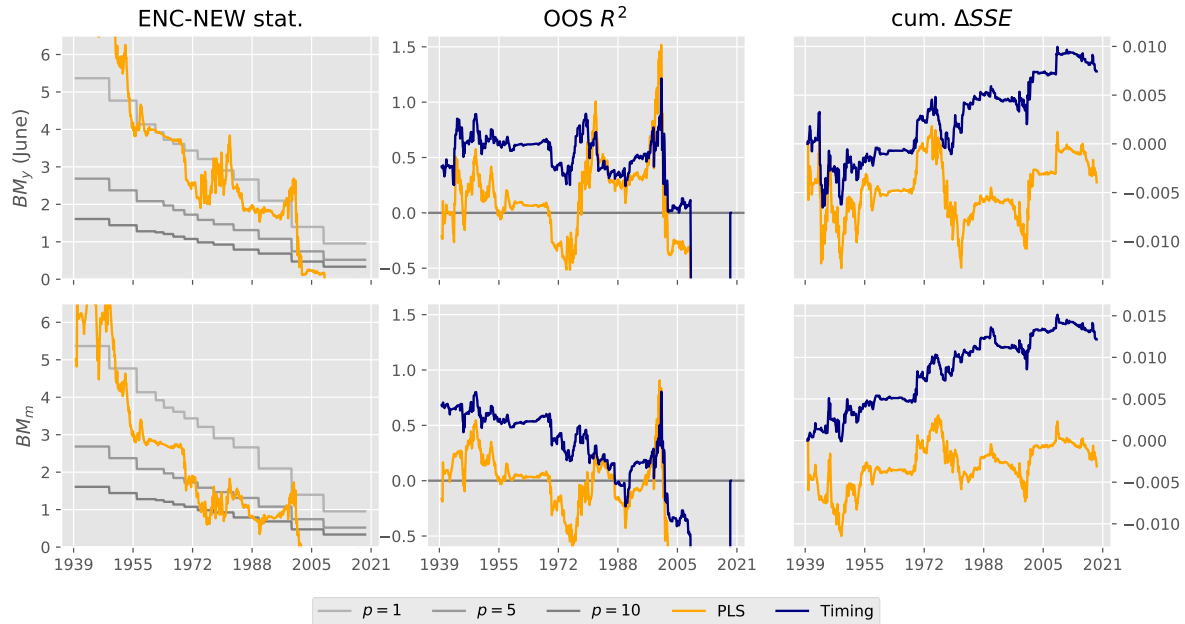
Figure 2: OOS recursive standard PLS and predictability timing PLS forecasts based on BMs of 6 BM/ME portfolios in January 1940–December 2019.

The graphs show (left) ENC-NEW test statistics of Clark and McCracken (2001), with gray kinked lines for the critical values at 1, 5, and 10 percent (from light to dark); (center) OOS R^2 by sample split date; and (right) the cumulative sum of squared forecasting errors of the historical mean minus the one from each model, ΔSSE. There are more extreme values outside the bounds in each graph (unreported for scaling reasons). (Light) orange is for the standard model, blue is for predictability timing.

Summary: The standard PLS model is already significant for most split dates, but predictability timing improves the performance of the forecasts and the OOS R^2 becomes positive since the Oil Shock, for example.

forecast months. All other performance measures are also substantially better in high forecast months: The Sharpe ratio in high (low) forecast months is 0.76 (-0.22), skewness is -0.43 (-0.84), and kurtosis is 3.64 (5.16).

In addition, the model predicts high premiums (in 318 months) around 55% more often than low premiums (in 203): This is a model of high premiums ("bad times") in the full sample. But the opposite happens after 1980. The remaining columns in Table 4a display the same statistics for the periods after 1980, 2000, or 2010 (columns $_{80+}$, $_{00+}$, and $_{10+}$). In all these periods, the model predicts high premiums less often than low premiums (the limit is 3/36, after 2010). The forecasts still identify differences in terms of returns,

Sharpe ratios, skewness, and kurtosis in all recent samples (except since 2010 for returns). However, the differentiation between high and low premiums is stronger in the full sample.

Table 4b is equivalent to Table 4a but displays statistics for average forecast months (by construction, months that follow realizations of inaccurate PLS forecasts), $MP^{neutral}$, and for all months in the respective sample, MP^{hist}. As ideally expected, returns and Sharpe ratios in average forecast months are lower (higher) than the ones observed in high (low) forecast months in the full sample: The mean return is 0.67% (2.90) and the Sharpe ratio is 0.41, between the extremes of low and high premiums displayed in Table 4a (but not robustly in subsamples). And a final interesting information in Table 4b is that the proportion of inaccurate PLS forecasts increases over time: From 601/1122 in the full sample to 68/107 since 2010: The accuracy of the PLS forecasts has deteriorated over time, in line with the results in de Oliveira Souza (2020b).

5   Summary

In this paper, we learn that a simple "predictability timing" procedure improves the OOS performance of the PLS forecasts of the equity premium in monthly frequency to the point of finally generating persistently positive OOS R^2, especially since the Oil Shock. This ultimately validates the model according to the hurdle in Welch and Goyal (2008).

We also learn that simple OLS forecasts based on the realized market premium as the only regressor produces a second valid market premium forecasting model, with large and positive OOS R^2 for almost every split date before 2010. This evidence supports the explanation in de Oliveira Souza (2020b) for the use of BMs with one extra lag as (PLS) regressors.

A Standard PLS performance

Table 5: Market premium after (standard) PLS forecasts above, MP^{high}, or below average, MP^{low} in July 1926 – December 2019 (or recently).

	(1) MP^{low}	(2) 80+	(3) 00+	(4) 10+	(5) MP^{high}	(6) 80+	(7) 00+
PLS_{μ}	0.43	0.44	0.35	0.86*	0.58**	0.76*	0.74
	(1.90)	(1.69)	(1.04)	(2.58)	(2.61)	(2.33)	(1.44)
SR_y	0.29	0.33	0.27	0.86	0.37	0.66	0.72
γ_1	-0.09	-1.36	-0.90	-0.55	-0.89	-0.02	-0.80
α_4	10.77	7.64	5.03	4.20	9.45	3.25	4.22
Months	524	319	179	107	597	148	48

Columns (1) to (4) are for months following below average forecasts, and (5) to (7) for above average forecasts. I repeat the evaluation of (1)/(5) in subsamples after 1980, 2000, and 2010 (shown as $_{80+}$, $_{00+}$, and $_{10+}$). PLS_{μ} is the mean return, OLS t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, SR_y is the annualized Sharpe ratio, γ_1 is skewness, α_4 is the kurtosis. PLS forecasts are based on monthly BMs with one extra lag. **Summary:** Full sample returns and Sharpe ratios are slightly larger in months predicted to have above average returns (but still lower than the ones when predictability timing forecasts are set to the mean in Table 4b, for example). Above average forecasts are the majority in the full sample (597/524), but decrease to zero after 2010.

References

Clark, Todd, and Michael McCracken, 2001, Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics* 105, 85–110.

Clark, Todd E., and Kenneth D. West, 2006, Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis, *Journal of Econometrics* 135, 155–186.

Cochrane, John H., 2017, Macro-finance, *Review of Finance* 21, 945–985.

Cujean, Julien, and Michael Hasler, 2017, Why does return predictability concentrate in bad times?, *Journal of Finance* 72, 2717–2758.

de Oliveira Souza, Thiago, 2020a, A critique of momentum anomalies, Unpublished working paper. University of Southern Denmark.

de Oliveira Souza, Thiago, 2020b, Dissecting market expectations in the cross-section of book-to-market ratios, Unpublished working paper. University of Southern Denmark.

de Oliveira Souza, Thiago, 2020c, Predictability concentrates in bad times. And so does disagreement, Unpublished working paper. University of Southern Denmark.

Kelly, Bryan, and Seth Pruitt, 2013, Market expectations in the cross-section of present values, *Journal of Finance* 68, 1721–1756.

Munk, Claus, 2013, *Financial asset pricing theory* (Oxford University Press).

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

Newey, Whitney K., and Kenneth D. West, 1994, Automatic Lag Selection in Covariance Matrix Estimation, *Review of Economic Studies* 61, 631–653.

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