Journal of Financial Markets 38 (2018) 83–102

Contents lists available at ScienceDirect

Journal of Financial Markets

journal homepage: www.elsevier.com/locate/finmar

Forecasting the equity risk premium: The importance of regime-dependent evaluation ☆

Nick Baltas [a,*], Dimitrios Karyampas [b]

[a] *Imperial College Business School, South Kensington Campus, SW7 2AZ London, UK*
[b] *Bocconi University, 20100 Milan, Italy*

A R T I C L E I N F O

Article history:
Received 6 December 2016
Received in revised form
10 November 2017
Accepted 12 November 2017
Available online 16 November 2017

JEL classification:
C53
C58
G11
G14
G17

Keywords:
Equity risk premium
Out-of-sample forecasting
Economic constraints
Predictive regression
Asset allocation
Business cycles

A B S T R A C T

Asset allocation is critically dependent on the ability to forecast the equity risk premium (ERP) out-of-sample. But, is superior econometric predictability across the business cycle synonymous with predictability at all times? We evaluate recently introduced ERP forecasting models, which have been shown to generate econometrically superior ERP forecasts, and find that their forecasting ability is regime-dependent. They give rise to significant relative losses during market downturns, when it matters the most for asset allocators to retain assets and their client base intact. Conversely, any economic benefit occurring during market upswings is diminished for high risk-averse and leverage-constrained investors.

1. Introduction

Asset allocation decisions, whether strategic or tactical, are closely related to forecasts of the equity risk premium (ERP). This renders the ERP forecasting model of critical importance for portfolio managers and asset allocators. Recently, a number of methodological advances have been introduced in the literature to improve the precision of ERP forecasts over a long sample period. But, is econometric predictability across the business cycle synonymous with greater predictability across different market regimes and most importantly during periods of market stress? Adding to the above, in making their investment decisions, investment professionals typically face a number of constraints, like short-selling or leverage constraints, which can potentially shrink the value add of a sophisticated ERP forecasting model.

☆ Comments by Stephen Brown, Mathijs Cosemans, Serge Darolles, Amit Goyal (the co-editor), Emmanuel Jurczenko (discussant), Simon Scheidegger, Guofu Zhou and an anonymous referee, as well as by participants at a seminar at EHL (Sept. 2016) and at the Computational and Financial Econometrics conference (Dec. 2015) are gratefully acknowledged. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The paper has been previously circulated with the title "Forecasting the Equity Risk Premium: Predictability versus Profitability."

* Corresponding author.
E-mail addresses: n.baltas@imperial.ac.uk (N. Baltas), dimitrios.karyampas@unibocconi.it (D. Karyampas).

https://doi.org/10.1016/j.finmar.2017.11.002
1386-4181/

The objective of this paper is twofold. Our first objective is to illustrate the importance of conditional performance evaluation of ERP forecasting models across different market regimes. In particular, we provide evidence that greater predictability over the business cycle does not necessarily imply predictability at all times. We study the conditional performance of ERP forecasting models and subsequently of the respective tactical asset allocation decisions separately across expansions and recessions, across up and down markets, and across high and low volatility regimes. Our regime-dependent analysis is justified by the fact that portfolio and asset managers are likely to exhibit asymmetric utility towards gains during good periods, as opposed to losses during bad periods; Petajisto (2013) argues that underperforming the benchmark is more painful during down markets than during up markets and for that reason a large proportion of active mutual funds have an incentive to remain closer to their benchmarks during down markets to avoid any outflows due to underperformance. Investors allocate more aggressively on the best-performing mutual funds (Sirri and Tufano, 1998; Huang et al., 2007), whereas poorly performing mutual funds are more likely to incur redemptions (O'Neal, 2004; Ivković and Weisbenner, 2009; Cashman et al., 2012). Similarly, hedge funds that do not produce alpha are more likely to liquidate (Fung et al., 2008). All this evidence suggests that the conditional performance of an ERP forecasting model across different market regimes is of critical importance for an asset allocator.

Our second objective is to investigate the effect that realistic investment constraints can have on the value-add of an ERP forecasting model for an investor who makes tactical asset allocation decisions based on the respective ERP forecasts of the model. Either by mandate or regulation, investors typically face a number of constraints, which put hard thresholds on minimum and maximum allocation across risky assets. A short-sale constraint does not allow the investor to take short positions, whereas a leverage constraint does not allow the investor to employ leverage. The tighter the constraints and/or the higher the risk aversion of the investor, the weaker the impact that a forecasting model can have in asset allocation decisions. Our aim is to carefully study these implications.

There are a number of challenges with forecasting excess returns in equity markets. Even though a long list of macroeconomic variables have been used to forecast equity market returns on an in-sample basis using a univariate linear ordinary least squares (OLS) framework,[1] they fail to generate more accurate ERP forecasts than simple ERP historical averages on an out-of-sample basis, as first highlighted by Goyal and Welch (2008). Timmermann (2008) finds that out-of-sample econometric predictability is typically episodic and short-lived -independent of the model or the forecasting variable (s) used- as investors who try to take advantage of it, act in a way that any predicted future return is instantly incorporated in the current price.[2]

We employ recent methodological refinements of the OLS forecasting model that have been introduced to provide econometrically more accurate out-of-sample ERP forecasts than historical means. Campbell and Thompson (2008) suggest truncating the ERP forecast to zero, if it is negative (the CT model, henceforth); the rationale being that market excess returns should be -in expectation- positive to justify equity investing. Based on the same economic reasoning, Pettenuzzo et al. (2014) introduce a Bayesian forecasting framework, under which conditional out-of-sample ERP forecasts are constrained to the positive territory (the PTV model, henceforth). Both models improve econometrically the out-of-sample forecasting ability of conventional macroeconomic predictors, at least unconditionally. However, in both papers the authors agree that the improvement is mainly due to the earlier years of their sample period, and not over the most recent decades, characterized by a number of substantial downturns, such as the collapse of the dot-com bubble, the credit crisis, and the eurozone crisis. This underperformance during turbulent periods motivates our analysis.

Following the above, we first question whether the improved econometric predictability of these constrained models actually improves the profitability of dynamic asset allocation plans, especially during market downturns. Given that these models generate non-negative and less volatile ERP forecasts, we hypothesize that they improve econometrically the accuracy only on an unconditional basis because the average ERP realization over the business cycle has historically been positive. Using a long sample period from January 1927 to December 2013 and the dividend-price ratio[3] we confirm that the constrained CT and PTV models improve the out-of-sample econometric accuracy of the unconstrained OLS model unconditionally (i.e., across the entire out-of-sample period). However, on a conditional basis, we find that these results are largely regime-dependent. Using a series of event studies across various market regimes, we provide strong statistical evidence that the constrained models are more accurate in up markets, during expansions, and during low-volatility periods. However, this completely reverses in down markets, during recessions, and during high-volatility periods with the constrained models generating substantially larger forecast errors than the OLS model.

What are the implications of this regime-dependent performance for a risk-averse investor who dynamically allocates between a risky asset (equity market) and a risk-free asset, using ERP forecasts of the various models under consideration? We conduct an asset allocation analysis and find that the constrained models generate positive certainty equivalent return

[1] For a long list of traditional ERP macroeconomic predictors see Goyal and Welch (2008), who are the first in the literature to highlight the importance of out-of-sample forecasting. Following Goyal and Welch (2008), new predictors have emerged, such as the output gap (Cooper and Priestley, 2009), the variance risk premium (Bollerslev et al., 2009), the nearness to the 52-week high and to the historical high (Li and Yu, 2012), the aggregate implied cost of capital (Li et al., 2013), price and volume based technical indicators (Neely et al., 2014), the end-of-year economic growth (Møller and Rangvid, 2015), an investor sentiment indicator (Huang et al., 2015), a mean-reversion indicator (Huang et al., 2017) and short interest (Rapach et al., 2016).

[2] This mechanism is in line with the adaptive markets hypothesis of Lo (2004).

[3] Ratios of dividends and (current or lagged) prices have been heavily used in the literature for ERP forecasting as the predictor of next month's S&P 500 Index excess returns, (e.g., Fama and French, 1988; Campbell and Shiller, 1988; Campbell and Yogo, 2006; Ang and Bekaert, 2007).

gains for the investor compared to the unconstrained OLS model, but these gains turn negative (so they turn into certainty equivalent return losses) during down and recessionary markets. Put differently, the constrained models generate pronounced performance drawdowns, and therefore their worst relative performance in periods when it becomes most important to deliver higher relative returns, either against the benchmark or the competition, so to retain assets and the client base intact. There is a significant relationship between under-/out-performance and capital out-/in-flows (Sirri and Tufano, 1998; O'Neal, 2004; Huang et al., 2007; Fung et al., 2008; Ivković and Weisbenner, 2009; Cashman et al., 2012).

We next examine the effect of constraints from a different perspective. In particular we investigate the value-add of the various ERP forecasting models under consideration, when the asset allocator faces different levels of constrained minimum and maximum equity allocation. For example, a short-sale constraint is -trivially- equivalent to a non-negative ERP forecast and therefore becomes redundant for constrained models, such as the CT and the PTV models. On the other hand, our analysis shows that a leverage constraint is more likely to become binding for the constrained models, as they typically generate larger ERP forecasts. This fact diminishes the value-add of the constrained models especially during up markets that are characterized by high ERP realizations. All in all, we find that the tighter the investment constraints become and the more risk averse the investor, the lower the value-add of a constrained ERP forecasting model. Put differently, whether a certain model improves (or not) econometrically the forecasting accuracy of ERP matters much less from a profitability point of view for a constrained and high risk-averse investor, than for an investor who can take short positions or most importantly employ leverage.

Taken together, our findings show that constrained ERP forecasting models can cause sizeable economic losses during market downturns, whereas their value-add during upswings is largely diminished if the investor faces leverage constraints. This evidence suggests that the greater econometric predictability of constrained models across the full sample does not necessarily imply better predictability at all times, especially when financial intermediation idiosyncrasies are taken into account. These findings have important implications for the design of new ERP forecasting frameworks and they clearly highlight the requirement for accommodating some form of regime dependency. Dangl and Halling (2012) suggest the introduction of time-variation in the coefficients of the linear forecasting model using a Bayesian framework in order to improve out-of-sample forecasting ability and find that this methodological amendment improves the performance of the model across both recessionary and expansionary periods. However, we consider the works by Zhu and Zhu (2013) and Huang et al. (2017) as being the first to explicitly highlight the need for designing forecasting models that can accurately forecast future ERP across both good and bad times. In addition, Tu (2010) shows that there are sizeable welfare losses associated with ignoring regime switching in asset allocation, which further demonstrates the need for such forecasting models.

The paper is organized as follows. Section 2 presents an overview of the ERP constrained and unconstrained forecasting models and of our dataset. Our empirical results on out-of-sample forecasting are presented in Section 3 and the respective asset allocation implications for a mean-variance investor are presented in Section 4. Finally, Section 5 concludes the paper.

2. Methodology

We start our analysis by providing an overview of the baseline unconstrained OLS forecasting model, of its recent constrained variants introduced by Campbell and Thompson (2008) and by Pettenuzzo et al. (2014), and of our dataset. We focus on out-of-sample ERP forecasting at the monthly horizon, as predictability across longer horizons can be an artefact of highly persistent predictor variables (Boudoukh et al., 2008; Cochrane, 2011). Longer-term ERP forecasting can be of practical importance for investors who have longer investment horizons, like pension funds. We have also analyzed quarterly and annual horizons and the results -available upon request- are in line with the findings presented in the paper.

2.1. Forecasting models

The baseline ERP forecasting model is a linear regression model under which the market excess return over the next period, r_{t+1}, is forecasted by the current value of a forecasting variable, x_t:

$$OLS: \quad \hat{r}^{OLS}_{t+1|t} = \hat{\alpha} + \hat{\beta}\cdot x_t, \tag{1}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are estimated using information up to time t from the OLS forecasting regression:

$$r_{\tau+1} = \alpha + \beta\cdot x_\tau + \epsilon_{\tau+1}, \quad \tau = 1, \ldots, t-1. \tag{2}$$

Despite the historical good in-sample performance of this model (i.e., using $\hat{\alpha}$ and $\hat{\beta}$ estimates from the entire sample) for a broad list of forecasting variables, its out-of-sample forecasting performance has been relatively poor, as shown in Goyal and Welch (2008). In an effort to improve its out-of-sample performance, Campbell and Thompson (2008) impose simple constraints on the OLS forecasts motivated by economic theory. In that respect, they suggest truncating the excess return forecast, $r^{OLS}_{t+1|t}$, if it is negative (they call this the “positive forecast” model):

$$CT:\quad \hat{r}^{CT}_{t+1|t} = max\left(\hat{\alpha} + \hat{\beta}\cdot x_t, 0\right). \tag{3}$$

The choice of the zero hard-coded lower bound in the ERP forecast is justified by the fact that market excess returns should be -in expectation- positive to justify equity investing. Campbell and Thompson (2008) show that the truncated model significantly improves the out-of-sample forecasting performance of a large list of predictor variables.

Following the same basic economic reasoning on the positivity of the forecasted ERP, Pettenuzzo et al. (2014) introduce a Bayesian forecasting framework, under which conditional out-of-sample excess return forecasts are constrained to a non-negative territory:

$$PTV:\quad \hat{r}^{PTV}_{t+1|t} = \bar{\alpha} + \bar{\beta}\cdot x_t. \tag{4}$$

where $\bar{\alpha}$ and $\bar{\beta}$ are the average intercept and slope of all pairs of values (α, β) that belong to a set $\mathcal{A}_t$ of admissible pairs that lead to positive in-sample and out-of-sample ERP forecasts:

$$\mathcal{A}_t = \{(\alpha, \beta): \alpha + \beta\cdot x_\tau \geq 0, \quad \forall\, \tau = 1, \ldots, t\} \tag{5}$$

The set $\mathcal{A}_t$ is formed at time t (i.e., at the end of each forecasting period) using the Gibbs sampler Bayesian estimation framework of PTV; for further details, we refer the reader to the Appendix of Pettenuzzo et al. (2014). The PTV model differs from the CT model in that it performs a constrained optimization in the actual fitting of the model as opposed to enforcing a post-fitting truncation in the ERP forecast.

In order to evaluate the forecasting ability of the various forecasting models we use the out-of-sample (OOS) R^2 statistic, R^2_{OOS}, which is defined as the proportional decrease in the mean squared forecast error (MSFE) between the model of interest and a benchmark model:

$$R^2_{OOS} = 1 - \frac{MSFE_{Model}}{MSFE_{Benchmark}}. \tag{6}$$

To the best of our knowledge, the historical average ERP is typically used as the benchmark model in the literature. However, as our objective is to compare the constrained CT and PTV models to the unconstrained OLS model, we estimate the R^2_{OOS} statistic using the unconstrained OLS model as the benchmark model and denote the statistic by $R^2_{OOS,OLS}$:

$$R^2_{OOS,OLS} = 1 - \frac{MSFE_{Model}}{MSFE_{OLS}} = 1 - \frac{\sum_{\tau=t_{OOS}}^{T}\left(r_\tau - \hat{r}^{Model}_{\tau|\tau-1}\right)^2}{\sum_{\tau=t_{OOS}}^{T}\left(r_\tau - \hat{r}^{OLS}_{\tau|\tau-1}\right)^2}, \tag{7}$$

where $Model = \{CT, PTV\}$, the time $\tau = t_{OOS}$ denotes the first month of out-of-sample forecasts, and T denotes the end of the sample period.

If $R^2_{OOS,OLS} > 0$, the constrained model is more accurate than the unconstrained OLS model, as it generates lower MSFE. Campbell and Thompson (2008) illustrate that a monthly R^2_{OOS} of 0.5% is enough to generate significant economic value for a mean-variance investor, who allocates between the equity market and a risk-free asset. We investigate these asset allocation implications in Section 4.

In order to statistically test the null hypothesis that the MSFE of the unconstrained OLS model is less than or equal to the MSFE of a constrained model (i.e. H_0: $R^2_{OOS,OLS} \leq 0$) against the one-sided, upper-tail, alternative hypothesis that the MSFE of the unconstrained OLS model is greater than the MSFE of a constrained model (i.e. H_A: $R^2_{OOS,OLS} > 0$) we use the Clark and West (2007) MSFE-adjusted statistic. This statistic is easily calculated by first forming the following variable across the out-of-sample period:

$$f_{t+1} = \left(r_{t+1} - \hat{r}^{OLS}_{t+1|t}\right)^2 - \left[\left(r_{t+1} - \hat{r}^{Model}_{t+1|t}\right)^2 - \left(\hat{r}^{OLS}_{t+1} - \hat{r}^{Model}_{t+1|t}\right)^2\right]. \tag{8}$$

The Clark and West (2007) MSFE-adjusted statistic is then the t-statistic from regressing the time series of f_{t+1} on a constant. A p-value for the one-sided, upper-tail test is conveniently obtained using the standard normal distribution.

2.2. Data description

Our sample period is from January 1927 to December 2013. The dependent variable in the forecasting models is the equity risk premium as proxied by the monthly excess total (i.e., including dividends) logarithmic return of the S&P 500 Index, as maintained by the Center of Research for Security Prices (CRSP). Over the sample period, the average annualized ERP has been 5.90% with a volatility of 19.15% (Sharpe ratio of 0.31) and a negative skewness of -0.42.

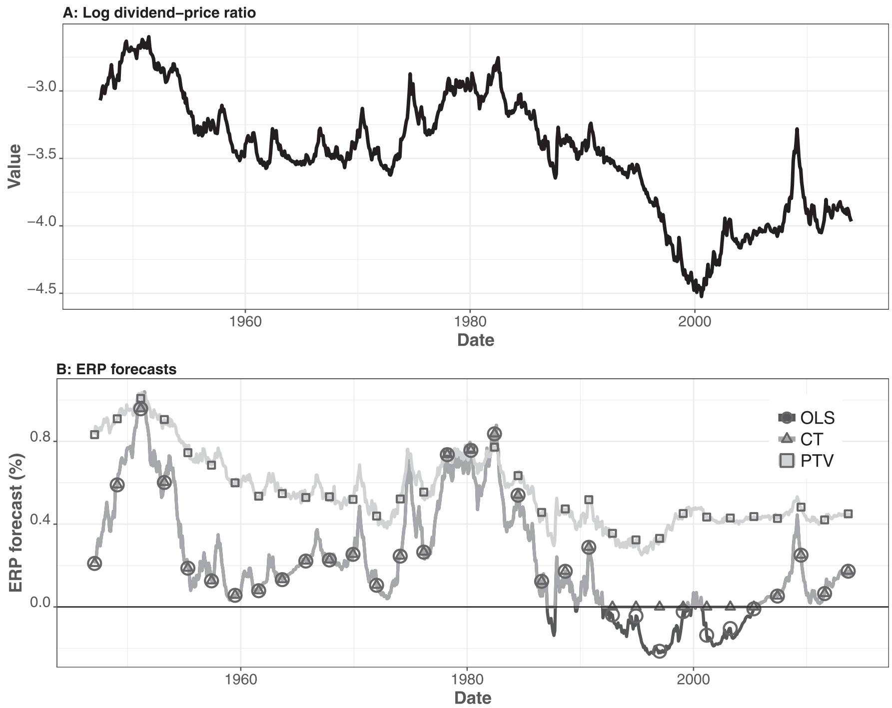

Fig. 1. ERP forecasts using the dividend-price ratio. The figure presents the historical values of the dividend-price ratio (difference between the log of the 12-month moving sums of dividends and the log of price of the S&P 500 Index) in Panel A and the ERP forecasts in Panel B. Three different forecasting models are used: the baseline OLS regression model, the CT truncation model, and the PTV Bayesian model. The initial training period is from January 1927 to December 1946 and the out-of-sample period is from January 1947 to December 2013.

Without loss of generality, we use the dividend-price ratio (d/p) of the S&P 500 Index as the predictor variable; the historical logarithmic values of d/p are shown in Panel A of Fig. 1. Monthly data for the S&P 500 Index and the d/p ratio are from the website of Amit Goyal.[4]

Our analysis does not constitute a comparison study between different predictor variables, but instead a comparison between the mechanics of different forecasting methodologies across various market regimes. For that purpose, our predictor (d/p) is inconsequential. In fact, our results remain both qualitatively and quantitatively robust for all the other predictor variables used by Goyal and Welch (2008).

3. Out-of-sample forecasting

Our first objective is to illustrate that superior econometric predictability over the business cycle is not necessarily synonymous to predictability across all market regimes. Along these lines, in this section we show that the constrained CT and PTV models are superior to the unconstrained model or to the historical ERP average on an unconditional basis (i.e., over the business cycle), and then focus more closely on the conditional performance of the various models across different market regimes.

For the empirical analysis, we follow the same setup as Goyal and Welch (2008) and Pettenuzzo et al. (2014) in that we reserve the first 20 years of the sample period for the initial training period of the forecasting models and then continue on an expanding window basis and generate monthly out-of-sample ERP forecasts for January 1947 onwards.[5]

Panel B of Fig. 1 presents the monthly out-of-sample ERP forecasts of the three models: the unconstrained OLS model, the CT truncation model, and the PTV Bayesian model.[6] The unconstrained OLS model generates both positive and negative ERP predictions, whereas the CT forecasts only differ from the OLS forecasts when the zero truncation constraint becomes

[4] See http://www.hec.unil.ch/agoyal/.

[5] Hansen and Timmermann (2012) provide guidance for an optimal split between in-sample and out-of-sample periods in order to avoid data mining in the documentation of statistical significant forecasting ability of a predictor when multiple sample split points are considered. However, we do not consider multiple split points and most importantly, we do not seek to document statistically strong ERP predictor variables. Instead, our objective is to compare the performance of various forecasting models across difference market regimes. This amounts to a horserace between different methodologies across market regimes, as opposed to a horserace between predictor variables over the entire out-of-sample period. As a result, the choices of the ERP predictor and of the initial split point between the in-sample and out-of-sample periods are both inconsequential in our paper.

[6] The PTV forecasts can be visually compared with Fig. 9 of Pettenuzzo et al. (2014).

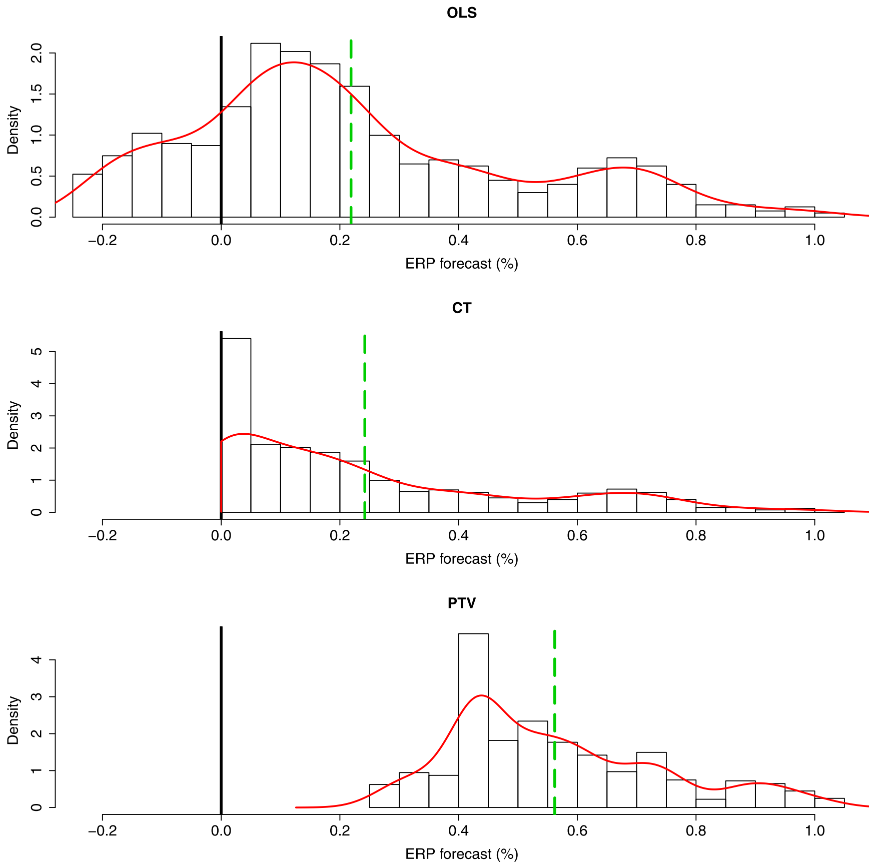

Fig. 2. Histograms of ERP forecasts. The figure presents the histograms of ERP forecasts using the dividend-price ratio as the predictor variable. Three different forecasting models are used: the baseline OLS regression model, the CT truncation model, and the PTV Bayesian model. The initial training period is from January 1927 to December 1946 and the out-of-sample period is from January 1947 to December 2013. The average forecast across the entire period for each predictor and forecasting model is indicated with a dashed vertical line.

binding; this is roughly during the 1992–2005 period. Contrary to OLS and CT forecasts, the PTV forecasts remain strictly positive across the entire sample period and exhibit relatively smaller time-series variation.

Fig. 2 presents the histograms of these out-of-sample ERP forecasts. Imposing progressively stricter constraints (first with the CT zero bound and subsequently with the strict positivity of PTV) naturally increases the average level of the ERP forecast and at the same time reduces the standard deviation around this average level.[7] This appears to come at odds with Huang et al. (2015), who argue that more volatile ERP forecasts are more likely to track more closely the largely volatile ERP realizations.

Given the dynamics of the different models, one can argue that the less constrained models are more likely to generate more accurate point forecasts in periods when the realization of the market excess return is negative, whereas the more constrained models are more likely to provide more accurate point forecasts when the realization of the market excess return happens to be positive. We can therefore hypothesize that the econometric superiority of the constrained models, as presented in the results of Campbell and Thompson (2008) and Pettenuzzo et al. (2014), is driven by the fact that the realized ERP has been on average positive over the long history. Numerically, the average ERP over the out-of-sample period (January 1947 to December 2013) has been 0.53% per month (6.4% per annum), which compared to the average OLS, CT, and PTV forecasts in Fig. 2 appears much closer to the constrained models than the unconstrained one.

3.1. *Forecasting accuracy against the historical average*

Before comparing the constrained forecasting models with the unconstrained OLS model, we first present how all these models compare against the historical ERP average forecast, which is the standard forecasting benchmark in the literature.

Following Campbell and Thompson (2008) and Goyal and Welch (2008), we compare the forecasting errors of all models,

[7] One can draw conceptual parallels to the work by Jagannathan and Ma (2003), who show that adding constraints when constructing a mean-variance efficient portfolio is equivalent to shrinking the covariance matrix.

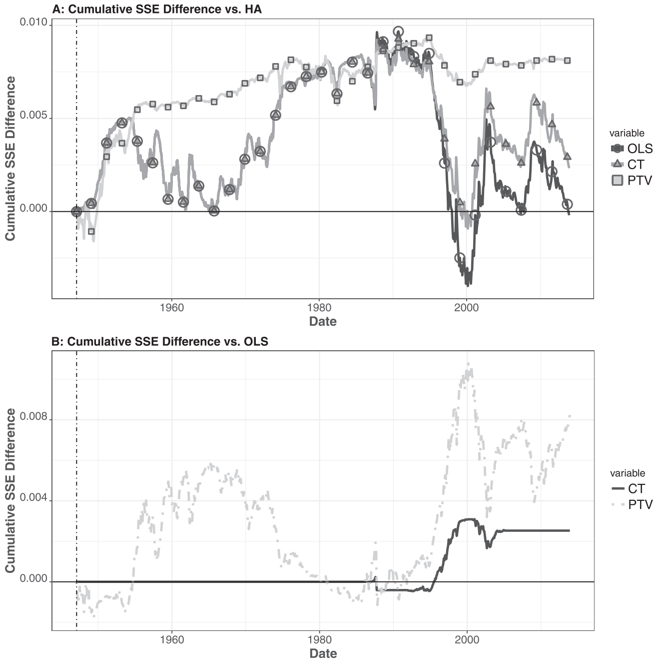

Fig. 3. Cumulative sum of squared forecast errors. The figure presents the difference in the cumulative sum of squared forecast errors between the three different forecasting models (the baseline OLS regression model, the CT truncation model, and the PTV Bayesian model) and the historical average (HA) in Panel A and the unconstrained OLS in Panel B. The predictor variable is the dividend-price ratio. The initial training period is from January 1927 to December 1946 and the out-of-sample period is from January 1947 to December 2013.

OLS, CT, and PTV, against the forecasting errors of the historical average (HA), which is similarly estimated on an expanding window basis using a 20-year initial training period. Panel A of Fig. 3 presents the difference in the cumulative sum of squared errors ($DCSSE$) between the OLS, CT, and PTV[8] models and the historical average:

$$DCSSE_{HA} = \sum_{\tau=t_{OOS}}^{T} \left(r_\tau - \hat{r}_{\tau|\tau-1}^{HA} \right)^2 - \sum_{\tau=t_{OOS}}^{T} \left(r_\tau - \hat{r}_{\tau|\tau-1}^{Model} \right)^2, \tag{9}$$

where $Model = \{OLS, CT, PTV\}$. An upward (downward) movement of the $DCSSE_{HA}$ indicates that the model is more (less) accurate in forecasting the ERP compared to the historical average. Trivially, if the $DCSSE_{HA}$ remains flat, the performance of the model of interest is similar to that of the historical average.[9]

The evidence in Panel A of Fig. 3 shows that the constrained models improve the forecasting accuracy of the unconstrained OLS model over the entire sample period, even though there are periods during which the latter generates significantly lower forecast errors. The CT model improves the performance of the OLS model following periods when the hard zero-bound constraint becomes binding (as identified in Fig. 2), in line with Campbell and Thompson (2008). The PTV model, which employs a stricter ERP constraint, generates relatively smoother $DCSSE_{HA}$ paths, but it seems that it achieves most of its forecasting outperformance during the very first years of the sample period; the $DCSSE_{HA}$ path increases strongly up until around 1956, then increases at a much lower pace up until 1995, before starting falling to around 2000 and remaining flat ever since.

To summarize, the constrained models appear to improve the forecasting power of the unconstrained model over the entire sample period, in line with Campbell and Thompson (2008) and Pettenuzzo et al. (2014), but this improvement tends to be achieved during earlier periods of the sample. The performance of the constrained models has been significantly challenged over the most recent decades. Seen from a different angle, this evidence seems to be in line with Timmermann (2008), who argues that individual forecasting models are likely to follow a life-cycle pattern, with their econometric accuracy falling after they become adopted by market participants.

3.2. Forecasting accuracy against the unconstrained OLS model

In column All in Table 1, we report the root mean square forecast error (RMSFE) of the three forecasting models (OLS, CT, and PTV), the $R^2_{OOS,OLS}$ statistic, and the proportion of months that a constrained model generates a smaller absolute forecast

[8] The results for the PTV model can be visually compared with Fig. 13 of Pettenuzzo et al. (2014).

[9] The use of the difference in the cumulative sum of squared errors as a tool of evaluating the forecasting performance of a model was first suggested by Goyal and Welch (2003).

Table 1
Out-of-sample performance.

		All	Down	Up
	# months	804	328	476
	OLS	4.227	4.746	3.828
RMSFE × 100	CT	4.223	4.763	3.806
	PTV	4.215	5.005	3.568
$R^2_{OOS,OLS}$	CT	0.18%**	−0.72%	1.12%***
	PTV	0.57%***	−11.25%	13.12%***
% of months with	CT	13.20%	0.30%	22.11%
smaller error vs. OLS	PTV	55.54%	1.52%	92.84%

This table presents the results of a comparison of the out-of-sample forecasting ability of the constrained CT and PTV models and the unconstrained OLS model when the dividend-price ratio is used as the predictor variable across the full sample ("All") and across months with a positive ERP realization ("Up") and months with a negative ERP realization ("Down"). The out-of-sample period is January 1947 to December 2013. The null hypothesis H_0: $R^2_{OOS,OLS} \leq 0$ is evaluated against the one-sided alternative hypothesis H_A: $R^2_{OOS,OLS} > 0$ based on the Clark and West (2007) MSFE-adjusted statistic; *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

than the unconstrained OLS model across the entire sample.[10] We find that both constrained models improve the forecasting ability of the unconstrained model in line with Campbell and Thompson (2008) and Pettenuzzo et al. (2014). The $R^2_{OOS,OLS}$ is 0.18% for the CT model and 0.57% for the PTV model, both statistically significant.

Panel B of Fig. 3 presents the *DCSSE* between the constrained models, CT and PTV, and the unconstrained OLS model:

$$DCSSE_{OLS} = \sum_{\tau=t_{OOS}}^{T} \left(r_\tau - \hat{r}^{OLS}_{\tau|\tau-1} \right)^2 - \sum_{\tau=t_{OOS}}^{T} \left(r_\tau - \hat{r}^{Model}_{\tau|\tau-1} \right)^2, \tag{10}$$

where $Model = \{CT, PTV\}$. This calculation resembles the conventional statistic of Eq. (9), but allows for the direct comparison between the unconstrained OLS model and its constrained variants. When the $DCSSE_{OLS}$ moves upwards (downwards) over time, the constrained model is more (less) accurate in forecasting the ERP compared to the unconstrained one.

The evidence is completely in line with our hypothesis. The constrained CT and PTV models improve econometrically the forecasting performance of the unconstrained OLS model on an unconditional basis (the respective $DCSSE_{OLS}$ lines end up in the positive territory at the end of the sample period and the $R^2_{OOS,OLS}$ values are all positive in Table 1), but there exist several periods during which the unconstrained model is more accurate.

The constrained models appear to outperform the baseline OLS model during market upturns, but significantly underperform during volatile periods of negative ERP realizations. Fig. 4 presents the 36-month rolling Sharpe ratio of the market. At times when the Sharpe ratio falls (due to either low ERP realizations or/and high market volatility) or, even worse, turns negative, the constrained forecasting models perform worse than the unconstrained model. Instead, at times when the Sharpe ratio increases, the $DCSSE_{OLS}$ of the constrained models follows closely.

The most recent decades (from the second half of 1990 s onwards) highlight this prescribed dependence between the performance of the constrained models and the ERP realizations even more clearly. A number of distinct periods such as (a) the build-up of the dot-com bubble in the late 1990 s with the subsequent collapse in the latter part of 2002, (b) the bull market of the following years, with the subsequent credit crisis in 2008, and (c) the most recent market rally that was temporarily hit by the eurozone crisis episodes in 2010 and 2011, are all patterns that are present in both the rolling Sharpe ratio calculation, and in the forecast errors of the constrained models.

To summarize, our full-sample results are in line with Campbell and Thompson (2008) and Pettenuzzo et al. (2014); the constrained forecasting models are superior to the unconstrained model or the historical average. However, the constrained forecasting models perform significantly better during up markets, whereas they underperform during periods with negative ERP realizations.

3.3. Forecasting accuracy and the market regime

The Down and Up columns of Table 1 present the various forecasting performance statistics for down and up markets,

[10] The value of the proportion of months that a constrained model generates a strictly smaller absolute forecast than the unconstrained OLS model should be treated with caution when it comes to the CT model. The CT truncation model generates the same ERP forecast as the OLS model, when the OLS-based ERP forecast is positive. As a result, for these periods the forecast error of the CT model is exactly the same as the OLS model. The value that we report identifies strictly lower absolute errors and does not account for equality in the errors. If we were to relax this strict condition and additionally allow for equality in the errors, the value for the CT model would largely increase. As an example, the estimates for the CT model in Table 1 would increase from 13.20% to 92.90%.

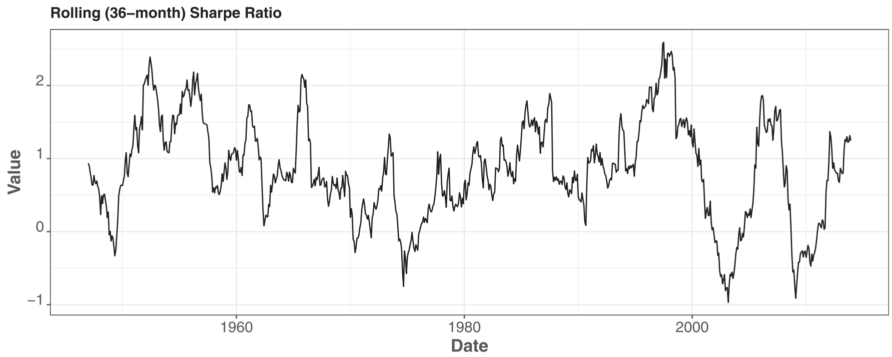

Fig. 4. Rolling Sharpe ratio of the S&P 500 Index. The figure presents the results of a rolling Sharpe ratio calculation at the end of each month for the S&P 500 Index using a window of 36 months. The sample period is from January 1947 to December 2013.

respectively, which are trivially defined by months with positive and negative ERP realization. Over our out-of-sample period, we identify 476 months with a positive ERP realization and 328 months with a negative ERP realization.

The empirical evidence is overwhelming. During up markets, the constrained models strongly outperform the unconstrained OLS model with $R^2_{OOS,OLS}$ values that are an order of magnitude larger than conventional out-of-sample R^2 estimates in the literature on ERP forecasting. The $R^2_{OOS,OLS}$ is 1.12% for the CT model and 13.12% for the PTV model, both statistically significant at the 1% level. The PTV model generates significantly larger $R^2_{OOS,OLS}$ values than the CT model, which shows that the stricter the constraint, the stronger the outperformance. On the contrary, during periods of negative ERP realizations, the constrained models exhibit significantly larger forecast errors, resulting in large, negative $R^2_{OOS,OLS}$, with the stricter PTV model suffering more than the CT model in complete symmetry to the up market result.

From an econometric perspective, it is important to infer whether the relative underperformance of the constrained models during down markets is statistically significant. To do so we use the Clark and West (2007) MSFE-adjusted statistic, but instead test for the null hypothesis H'_0: $R^2_{OOS,OLS} \geq 0$ against the one-sided, lower-tail, alternative hypothesis H'_A: $R^2_{OOS,OLS} < 0$; notice that this is the complete opposite statistical test than the one reported in all our tables, hence the use of the prime in the notation of the null and alternative hypotheses. In order to avoid confusion, we do not report the statistical significance of the $R^2_{OOS,OLS}$ values during down markets in Table 1, based on this framework. Instead, it suffices to say that H'_0 is strongly rejected across all down market instances, for both constrained CT and PTV models; the rejection is at 1% level.[11]

In order to visually present this state-dependent behaviour of the various models, in Fig. 5 we extend Panel B of Fig. 3 by splitting the $DCSSE_{OLS}$ calculation across months with positive EPR realization ($DCSSE^+_{OLS}$) and months with negative ERP realization ($DCSSE^-_{OLS}$):

$$DCSSE^+_{OLS} = \sum_{\tau=t_{OOS}}^{T} \left(r_\tau - \hat{r}^{OLS}_{\tau|\tau-1}\right)^2 \cdot \mathcal{I}_{r_\tau \geq 0} - \sum_{\tau=t_{OOS}}^{T} \left(r_\tau - \hat{r}^{Model}_{\tau|\tau-1}\right)^2 \cdot \mathcal{I}_{r_\tau \geq 0} \tag{11}$$

$$DCSSE^-_{OLS} = \sum_{\tau=t_{OOS}}^{T} \left(r_\tau - \hat{r}^{OLS}_{\tau|\tau-1}\right)^2 \cdot \mathcal{I}_{r_\tau < 0} - \sum_{\tau=t_{OOS}}^{T} \left(r_\tau - \hat{r}^{Model}_{\tau|\tau-1}\right)^2 \cdot \mathcal{I}_{r_\tau < 0} \tag{12}$$

where $\mathcal{I}_{r_\tau \geq 0}$ and $\mathcal{I}_{r_\tau < 0}$ denote indicator functions for positive and negative ERP realizations.

The evidence in Fig. 5 is strong and justifies the magnitude of the $R^2_{OOS,OLS}$ values in Table 1. By conditioning the evaluation of the forecasting performance of the models on the sign of the ERP realization, we confirm our hypothesis that the unconstrained OLS model outperforms the constrained models in months of negative ERP realization, whereas the constrained models outperform the unconstrained OLS model in months of positive ERP realization. Put differently, failing to forecast -even the sign of- the direction of market returns can have important implications when converting any forecast signals into a trading strategy; we focus on these issues in the next section. For completeness and even though it is out of the scope of this paper, it should be noted that forecasting the direction of the market (e.g., Henriksson and Merton, 1981; Pesaran and Timmermann, 1992) has been shown to be more closely related to a profitable trading strategy (Leitch and Tanner, 1991).

In order to further elaborate on these results, we next conduct a more granular analysis and focus not only on the sign of the ERP realization, but also on the magnitude of the realized return. We therefore divide the group of months with a negative ERP realization into negative-large (NL) and negative-small (NS) subsets, and equivalently divide the group of months with a positive ERP realization into positive-large (PL) and positive-small (PS) subsets. Table 2 presents the same forecasting performance results as in Table 1 across all four regimes.

[11] Clark and West (2007) MSFE-adjusted statistic values are available upon request from the authors.

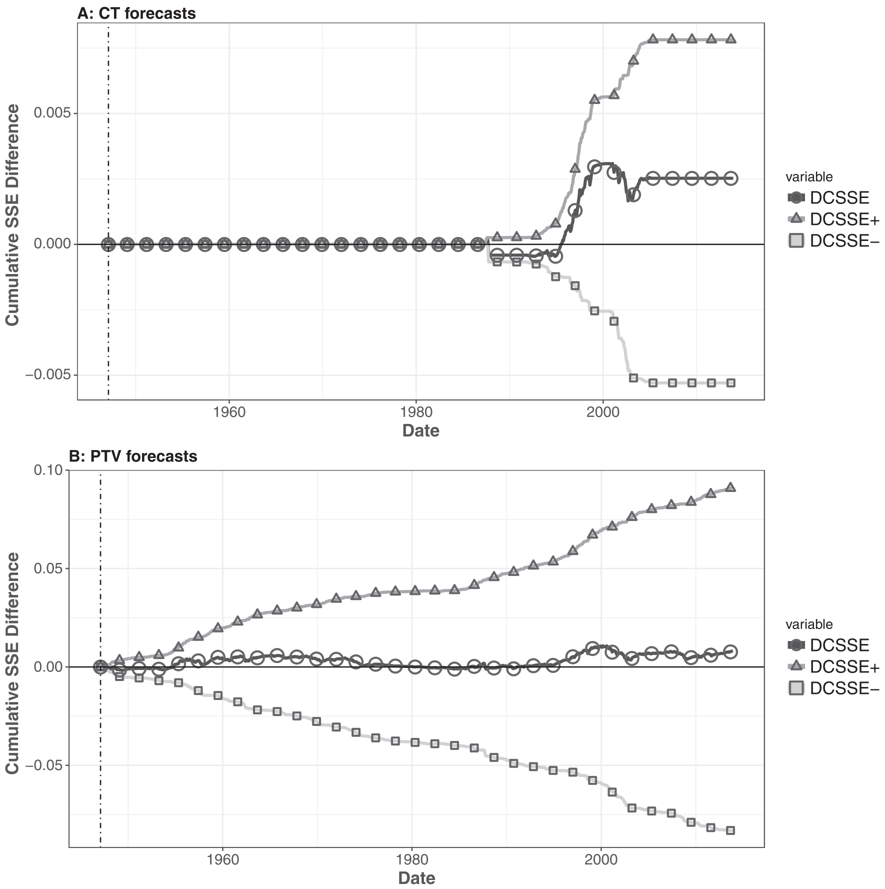

Fig. 5. Constrained versus unconstrained models in up and down markets. The figure presents the difference in the cumulative sum of squared forecast errors (*DCSSE*) between the unconstrained OLS model and the constrained models; the CT truncation model in Panel A and the PTV Bayesian model in Panel B. The predictor variable is the dividend-price ratio. The initial training period is from January 1927 to December 1946 and the out-of-sample period is from January 1947 to December 2013.

Table 2
Forecasting across market regimes.

		NL	NS	PS	PL
	# Months	164	164	238	238
ERP Realizations	Min	−25.0%	−2.4%	0.0%	2.9%
	Max	−2.4%	0.0%	2.9%	15.0%
RMSFE × 100	OLS	6.512	1.625	1.476	5.213
	CT	6.533	1.638	1.451	5.188
	PTV	6.813	1.921	1.187	4.909
$R^2_{OOS,OLS}$	CT	−0.66%	−1.64%	3.37%***	0.94%***
	PTV	−9.47%	−39.71%	35.33%***	11.33%***
% of months with Smaller error vs. OLS	CT	0.00%	0.61%	23.95%	20.25%
	PTV	2.44%	0.61%	88.24%	97.47%

The table presents the results of an event study comparing the out-of-sample forecasting ability of the constrained CT and PTV models and the unconstrained OLS model when the dividend-price ratio is used as the predictor, across four market regimes. The market regimes are determined by first dividing the entire out-of-sample period (804 months in total) between months with a positive ERP realization (476 months) and months with a negative EPR realization (328 months). Then, the negative ERP return bucket is equally divided into negative-large (NL) and negative-small (NS) subsets, and equivalently the positive ERP return bucket is equally divide into positive-large (PL) and positive-small (PS) subsets. The table reports the ERP boundaries (min/max) between the various market regimes. The out-of-sample period is from January 1947 to December 2013. The null hypothesis H_0: $R^2_{OOS,OLS} \leq 0$ is evaluated against the one-sided alternative hypothesis H_A: $R^2_{OOS,OLS} > 0$ based on the Clark and West (2007) MSFE-adjusted statistic; *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3
ERP forecasting across business cycles and volatility regimes.

		Business cycles		Volatility regimes		
		Rec	Exp	Low	Med	High
	# Months	122	682	268	268	268
RMSFE × 100	OLS	5.774	3.885	2.848	3.609	5.696
	CT	5.777	3.879	2.838	3.602	5.697
	PTV	5.855	3.848	2.702	3.587	5.753
$R^2_{OOS,OLS}$	CT	−0.13%	0.30%***	0.70%***	0.39%***	−0.04%
	PTV	−2.82%	1.92%***	9.98%***	1.23%**	−2.04%
% of months with Smaller error vs. OLS	CT	2.46%	15.12%	11.19%	13.11%	15.30%
	PTV	43.44%	57.71%	66.04%	54.68%	45.90%

The table presents the results of an event study comparing the out-of-sample forecasting ability of the constrained CT and PTV models and the unconstrained OLS model when the dividend-price ratio is used as the predictor variable, across recessionary (Rec) and expansionary (Exp) periods, as determined by the National Bureau of Economic Research (NBER) and across three volatility regimes: low, medium, high. The grouping of months in volatility regimes is done based on the monthly realized volatility of S&P 500 Index (sum of daily logarithmic returns). The out-of-sample period is January 1947 to December 2013. The null hypothesis H_0: $R^2_{OOS,OLS} \leq 0$ is evaluated against the one-sided alternative hypothesis H_A: $R^2_{OOS,OLS} > 0$ based on the Clark and West (2007) MSFE-adjusted statistic; *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

In line with our findings, the results in Table 2 shows that the constrained models strongly outperform the unconstrained OLS model over months with positive ERP realizations (PL and PS), but also strongly underperform the unconstrained OLS model over months with negative ERP realizations (NL and NS).

The constrained models appear to perform best, relative to the unconstrained model, in the PS regime, when the ERP is positive and small in the 0–2.9% range. The $R^2_{OOS,OLS}$ reaches extremely large and statistically significant values, all at the 1% level. Focusing on the most constrained PTV model, the $R^2_{OOS,OLS}$ is 35.33%; this effectively amounts to a reduction of the MSFE of the unconstrained OLS model by more than one third. Regarding the PL regime, when the ERP realizations range from 2.9% to 15%, the forecasting errors of all models increase uniformly as indicated by the RMSFE values. The constrained models still significantly outperform the unconstrained OLS model, but the relative reduction of the forecast errors, as captured by the $R^2_{OOS,OLS}$, is not as large as it is for the PS regime. However, the benefit from using constrained ERP forecasts is still pronounced, with all the $R^2_{OOS,OLS}$ values in the PL regime being statistically significant at the 1% level. The PTV model exhibits $R^2_{OOS,OLS}$ values of 11.33%.

Contrary to the positive ERP regimes, the constrained models do not perform similarly in the NS and NL regimes. In complete symmetry to the PS and PL regimes, the underperformance compared to the unconstrained OLS model is relatively larger in the NS regime, when the ERP is small and negative, in the -2.4% to 0% range. The $R^2_{OOS,OLS}$ of the PTV model is -39.71%; the respective values for the CT model are relatively lower, but still negative. This corroborates our earlier findings regarding the very poor performance of the constrained models in periods of negative ERP realizations. Importantly, the fact that the constrained models perform much worse than the unconstrained OLS model during the NS regime, can turn out to be detrimental for an investor who allocates between the equity market and a risk-free asset, as she could potentially suffer large relative losses if the market realizes even a small negative return. We leave the discussion of the asset allocation implications of the various forecasting models for Section 4.

As a last piece of evidence in this market regime analysis, Table 3 reports forecasting performance results separately for recessionary and expansionary periods as determined by the National Bureau of Economic Research (NBER).[12]

The empirical evidence is again completely in line with our hypothesis. The constrained models significantly outperform the constrained OLS model during expansionary periods (682 months in our sample), with the respective $R^2_{OOS,OLS}$ being positive and statistically significant at the 1% level. For recessionary periods (122 months in our sample), the unconstrained OLS model provides more accurate ERP forecasts, which can have important asset allocations implications.[13]

[12] The analysis across recessionary and expansionary periods has been relatively common in the ERP forecasting literature; see Rapach et al. (2010), Henkel et al. (2011), Neely et al. (2014), Pettenuzzo et al. (2014), and Huang et al. (2015, 2017).

[13] Interestingly enough, a large number of papers that introduce methodological improvements for ERP forecasting (e.g., Rapach et al., 2010; Henkel et al., 2011; Dangl and Halling, 2012; Zhu and Zhu, 2013; Neely et al., 2014), find that the forecasting ability of their models is only statistically strong (or at the very least, stronger) during down markets and recessionary periods. This is because they do not employ non-negative forecast restrictions as Campbell and Thompson (2008) and Pettenuzzo et al. (2014) do and as a consequence they do not incur the effects that we demonstrate.

3.4. Forecasting accuracy and the volatility regime

Equity returns and volatility innovations are typically negatively correlated (e.g., Ang and Bekaert, 2002) with spikes in volatility occurring mostly during bad periods with negative market returns. It is therefore worth exploring the forecasting ability of the various ERP forecasting models across different volatility regimes. Forming expectations, we conjecture that high volatility regimes would hurt the performance of constrained models, because such periods are more likely to experience negative return realizations. Instead, during periods of milder volatility, the constrained models are expected to outperform the unconstrained model.

Table 3 presents the results of the comparison of the forecasting ability of the various ERP forecasting models across three volatility regimes. In particular, using monthly estimates of market volatility (sum of daily squared logarithmic returns of the S&P 500 Index), we split the out-of-sample period into months of low volatility, medium volatility, and high volatility; the breakpoints between the regimes are 9.71% between low and medium volatility and 13.92% between medium and high volatility. These breakpoints are largely dependent on our sample period and are bound to change if one uses a different one. Our results should be regarded as an in-sample study of the dynamics of the various forecasting models across different volatilities environments.

The empirical evidence shows that the constrained CT and PTV models generate statistically lower forecast errors than the unconstrained OLS model in periods of low and medium volatility; as an example, the $R^2_{OOS,OLS}$ of the PTV model in the low-volatility regime is 9.98%, and is statistically strong at the 1% level. Conversely, during periods of high volatility, the constrained models generate substantially larger forecast errors compared to the unconstrained OLS model.

3.5. Robustness analysis: alternative CT models

In their study, Campbell and Thompson (2008) suggest a number of other constrained ERP forecasting models, apart from the truncation model, which they call "positive forecast." For robustness purposes, we subsequently evaluate the performance of two additional specifications: "positive slope", "fixed coefs."

In the positive slope model, the regression coefficient in Eq. (1) is set equal to zero if it has the "wrong" sign (as determined by the "theoretically expected sign" using a full-sample regression). Over our out-of-sample period, January 1947 to December 2013, the slope coefficient of the d/p ratio has always been positive, ranging between 0.26 and 0.53.[14] Along these lines, the positive slope model for the d/p ratio is identical to the unconstrained OLS model for our sample period. This is in line with Campbell and Thompson (2008), who find that the restriction on the slope coefficient of the d/p ratio is only binding in the 1930s; their out-of-sample period starts in 1927, unlike ours which starts in 1947.[15]

In the fixed coefs model, the restrictions of the steady-state theory are imposed and the intercept of the OLS regression is restricted to be equal to zero and the slope coefficient to be equal to one. Such a constraint effectively amounts to assuming that the ERP forecast over the next month is identical to last month's d/p ratio. Given that the d/p ratio is by construction positive,[16] the ERP forecasts of the fixed coefs model are always positive (see the logarithmic values of d/p ratio in Panel A of Fig. 1). As a result, the fixed coefs model is in line with the non-negativity design principle of the other constrained forecasting models (CT, PTV); in fact, notice that stricter constraints for the ERP forecasts are imposed by the fixed coefs model compared to the CT truncation model, as the former does not practically allow for forecasts equal to zero. Campbell and Thompson (2008) find that the fixed coefs model performs the best among the constrained models that they test for the d/p ratio; they report R^2_{OOS} of 0.42% versus –0.66% for the unconstrained OLS model.[17] We next assess its performance relative to the CT positive forecast and the PTV models to verify these findings.

Table 4 presents the results in terms of $R^2_{OOS,OLS}$ for the fixed coefs model across the various market regimes. For comparability purposes, we add our results for the CT truncation (positive forecast) and the PTV models. In line with Campbell and Thompson (2008), the fixed coefs model produces a higher $R^2_{OOS,OLS}$ than the CT truncation model over the full sample period. However, when conducting regime-dependent performance evaluation, we find that the fixed coefs model performs better than the CT truncation model in up markets, expansions, and low volatility regimes. Conversely, it performs worse in down markets, recessions, and high volatility regimes. These findings strongly corroborate our earlier findings and are in line with our main hypothesis: the out- and under-performance of the fixed coefs model versus the CT truncation model across market regimes aligns perfectly with their relative strictness on the ERP forecasts that they produce.

[14] We use the logarithmic d/p ratio as the predictor variable, following Goyal and Welch (2008). For the analysis of the positive slope model, we use the actual level of the d/p ratio to match the analysis by Campbell and Thompson (2008). Overall, whether one uses the actual d/p ratio or its logarithmic value does not alter qualitatively the results. The slope coefficient over our sample period is positive, as theoretically expected.

[15] Campbell and Thompson (2008) also combine the positive slope and positive forecast models in what they trivially call the "both" model. Given that in our sample period, the positive slope is essentially identical to the unconstrained OLS model, it becomes obvious that the both model degenerates into the positive forecast model (what we call the CT truncation model) that we have already studied in previous subsections.

[16] Technically, the d/p ratio can become equal to zero for a non-dividend paying stock. However, at the index level, as in our case, there is no month in our sample period where the d/p ratio was equal to zero.

[17] Notice that these estimates of R^2_{OOS} reported by Campbell and Thompson (2008) assume the historical average as the benchmark model; see our Eq. (6). As a result, they are not directly comparable to the $R^2_{OOS,OLS}$ estimates of our Table 4.

Table 4
ERP forecasting: robustness analysis.

	All	Down	Up	Business cycles		Volatility regimes		
				Rec	Exp	Low	Med	High
CT-Pos. Forecast	0.18%**	−0.72%	1.12%***	−0.13%	0.30%***	0.70%***	0.39%***	−0.04%
CT-Fixed Coefs	0.28%*	−1.84%	2.53%***	−0.59%	0.62%***	2.56%***	1.19%***	−0.66%
PTV	0.57%***	−11.25%	13.12%***	−2.82%	1.92%***	9.98%***	1.23%**	−2.04%

The table presents the out-of-sample (OOS) R^2 with respect to the unconstrained OLS model for the constrained CT positive forecast, fixed coefs, and PTV models across various market regimes, when the dividend-price ratio is used as the predictor variable. The out-of-sample period is January 1947 to December 2013. The null hypothesis H_0: $R^2_{OOS,OLS} \le 0$ is evaluated against the one-sided alternative hypothesis H_A: $R^2_{OOS,OLS} > 0$ based on the Clark and West (2007) MSFE-adjusted statistic; ∗, ∗∗, and ∗∗∗ denote significance at the 10%, 5%, and 1% levels, respectively.

When comparing the fixed coefs model to the PTV model, we find that the PTV model outperforms fixed coefs in the full sample, as well as during expansionary and low volatility periods; the performance is reversed when we focus on recessions and high volatility regimes with PTV producing worse $R^2_{OOS,OLS}$ values than either of the CT or the fixed coefs model.

In summary, our results are robust when the fixed coefs model is taken into account. Importantly, as the level of strictness on the ERP forecasts progressively increases in the OLS, CT truncation, fixed coefs, and PTV models, the outperformance improves over the entire sample period, as well as during expansionary and low volatility periods, but worsens significantly when the equity markets perform poorly, that is, during recessionary and high volatility periods. What are the implications for an asset allocator? As these constrained forecasting models struggle across down markets, recessions, and/or high volatility periods, it is reasonable to argue that they constitute a substantial challenge for the actual profitability of an investment strategy that is based on the ERP forecasts of these models.

4. Asset allocation implications

In this section, we focus on the implications of the regime-dependent performance of the various ERP forecasting models for an asset allocator who uses these ERP forecasts in order to tactically allocate between the equity market and a risk-free asset.[18] Additionally, we investigate how realistic investment constraints, like short-selling or leverage constraints, can affect the profitability of such a dynamic trading strategy.

Assuming mean-variance preferences, an investor with relative risk aversion γ would use an out-of-sample ERP forecast at the end of each month, in order to decide upon the amount of wealth to be invested in the equity market over the course of the following month:

$$w_t = \frac{1}{\gamma} \cdot \frac{\hat{r}_{t+1|t}}{\hat{\sigma}^2_{t+1|t}}, \tag{13}$$

where $\hat{r}_{t+1|t}$ denotes the out-of-sample ERP forecast based on a forecasting model and on information up until the end of month t and $\hat{\sigma}^2_{t+1|t}$ denotes the ERP forecast variance. We assume that the investor uses a rolling five-year historical estimate for the ERP variance (the same approach is followed by Campbell and Thompson (2008) and Huang et al. (2015, 2017)).

The large majority of studies that use the same methodology in order to evaluate the economic benefit of ERP forecasting models employ lower and maximum weights in an effort to limit the amount of shorting and leverage that is employed. Along these lines, the equity weight is bounded below and above by discretionary lower, w_{min}, and upper, w_{max}, bounds:

$$w_t = \max\left\{ \min\left[\frac{1}{\gamma} \cdot \frac{\hat{r}_{t+1|t}}{\hat{\sigma}^2_{t+1|t}}, w_{max} \right], w_{min} \right\}. \tag{14}$$

In our baseline case, we assume a risk aversion equal to three, a minimum weight of -0.5 (i.e., we allow for shorting up to 50%), and a maximum weight of 1 (i.e., we do not allow for leverage); Huang et al. (2017) use the same set of parameter values. Different levels of w_{min} and w_{max}, as well as of the risk aversion parameter have significant implications in the performance of the portfolio.

The portfolio return, $r_{p,t+1}$, is a linear combination of next month's equity market excess return, r_{t+1}, and the risk-free rate, $r_{f,t+1}$, which prevails at the end of month t:

$$r_{p,t+1} = w_t \cdot r_{t+1} + r_{f,t+1}. \tag{15}$$

[18] The importance of relating the econometric predictability of stock returns to economic profitability of an asset allocation framework was first discussed by Kandel and Stambaugh (1996). See also the more recent work by Timmermann (2008) and the follow-up commentary by Brown (2008).

Table 5
Asset allocation results.

Forecast Model	CER gain	Volatility	Sharpe ratio
Panel A: Overall			
CT	0.52%	8.82%	0.79
PTV	1.80%	12.86%	0.74
Panel B: Down markets			
CT	−1.66%	7.38%	−1.84
PTV	−19.02%	9.56%	−3.18
Panel C: Up markets			
CT	2.16%	7.28%	2.91
PTV	18.02%	7.82%	4.76
Panel D: Recessions			
CT	−0.33%	15.76%	0.06
PTV	−6.70%	18.88%	−0.20
Panel E: Expansions			
CT	0.67%	6.86%	1.17
PTV	3.34%	11.34%	1.05

The table reports the economic benefit for a mean-variance investor with a risk aversion of three ($\gamma = 3$) that uses constrained CT and PTV models in order to generate ERP forecasts using the dividend-price ratio as the predictor variable. The reported statistics are: the certainty equivalent returns (CER) gain against using an unconstrained OLS model, the annualized volatility of the portfolio between the equity market and the risk-free asset and the Sharpe ratio of the portfolio, defined as the monthly portfolio return in excess of the risk-free rate divided by the volatility of the portfolio return. The equity weight is lower bounded by $w_{min} = -0.5$ (allows for shorting of up to 50%) and upper bounded by $w_{max} = 1$ (no leverage allowed). The ERP forecast variance is estimated on a rolling basis using a five-year window. The out-of-sample period is January 1947 to December 2013.

The economic benefit of the various ERP forecasting models is measured by the means of the certainty equivalent return (CER), which is defined as the spread between the average portfolio return, $\hat{\mu}_P$, and its respective variance, $\hat{\sigma}_P^2$, scaled by 0.5 times the level of risk aversion:

$$CER = \hat{\mu}_P - \frac{\gamma}{2} \cdot \hat{\sigma}_P^2. \tag{16}$$

The CER represents the risk-free return that would render the mean-variance investor indifferent to investing in the proposed strategy. Alternatively, the CER can also be interpreted as the annual fee that the investor would be willing to pay to have access to the respective ERP forecast and therefore to exploit it. For that reason, all CER estimates in our analysis are multiplied by 12 to represent annual estimates.

We quantify the relative economic benefit by reporting the CER gain of a constrained model versus the unconstrained OLS model:

$$CER\ gain\ vs.\ \ OLS = CER_{Model} - CER_{OLS}, \tag{17}$$

where $Model = \{CT, PTV\}$. Notice that the conventional definition of the CER gain in the literature is with respect to the CER of a strategy that uses the historical average return as the ERP forecast. Our definition uses instead the unconstrained OLS model as the benchmark forecasting model, because we investigate the added economic benefit of a constrained model relative to an unconstrained one.

Table 5 reports the CER gain, the volatility, and the Sharpe ratio of the portfolio when using constrained CT and PTV forecasts for our baseline set of parameters ($\gamma = 3$, $w_{min} = -0.5$, $w_{max} = 1$) across the entire out-of-sample period, as well as across a number of market regimes. The results are in line with our hypothesis and empirical findings so far. The constrained forecasting models generate positive CER gains against the unconstrained OLS model across the entire sample period (Panel A), which shows that these models generate larger economic benefits for a risk-averse investor across the business cycle. However, when studying the performance of the models conditional on the market regime (down and up markets in Panels B-C and NBER expansions and recessions in Panels D-E), the CER gains of the constrained model appear to largely accumulate during up markets and expansionary periods. Down markets and recessionary periods are typically accompanied by negative CER gains (or, equivalently, the unconstrained OLS model generates CER gains against its constrained variants).

These results illustrate that a trading strategy that is based on the theoretically more accurate constrained ERP forecasting models can give rise to large relative losses during down and volatile markets, which has significant implications for the fund management industry. Investors tend to allocate more aggressively into the best-performing mutual funds (Sirri and Tufano, 1998; Huang et al., 2007), whereas mutual funds that perform poorly relative to their peers are more likely to incur redemptions (O'Neal, 2004; Ivković and Weisbenner, 2009; Cashman et al., 2012). Similarly, hedge funds that do not produce alpha are more likely to liquidate (Fung et al., 2008). Furthermore, Petajisto (2013) shows that active mutual fund managers prefer to stay close to their respective benchmarks during down markets, because underperforming the

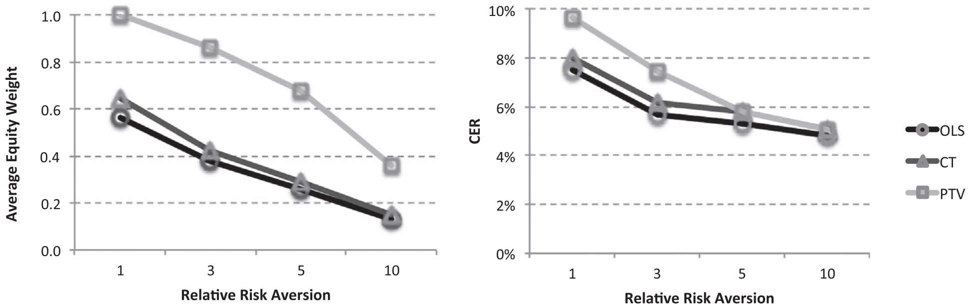

Fig. 6. The effect of risk aversion. The figure presents the marginal impact on the average equity weight (left diagram) and on the CER (right diagram) that is implied for different levels of relative risk aversion, when using out-of-sample ERP forecasts that are generated by the dividend-price ratio using the unconstrained OLS model, the constrained CT model, and the constrained PTV model. The minimum and maximum weights are fixed for all the scenarios to $w_{min} = -0.5$ and $w_{max} = 1$ respectively. The out-of-sample period is from January 1947 to December 2013.

benchmark is more painful during down markets and leads to outflows. Seen from a different perspective, these results are consistent with Tu (2010), who shows that there are sizeable certainty equivalent losses associated with ignoring regime switching in asset allocation.

We next evaluate the marginal effects of the three main parameters of the asset allocation experiment: the level of risk aversion, the minimum equity weight, w_{min}, and the maximum equity weight, w_{max}. To put our results in perspective, Campbell and Thompson (2008), Goyal and Welch (2008), Rapach et al. (2010), Neely et al. (2014), and Huang et al. (2015, 2017) assume $w_{min} = 0$ and $w_{max} = 1.5$; Rapach et al. (2016) assume $w_{min} = -0.5$ and $w_{max} = 1.5$, whereas Ferreira and Santa-Clara (2011) and Pettenuzzo et al. (2014) do not impose any restrictions. The typical level of relative risk aversion that is used in these studies is $\gamma = 3$; Ferreira and Santa-Clara (2011) use $\gamma = 2$, Pettenuzzo et al. (2014) present results for $\gamma = 2, 5, 10$, and Huang et al. (2015) present results for $\gamma = 1, 3, 5$.

4.1. The effect of risk aversion

Following from Eq. (14), the higher the risk aversion, the lower the level of equity investing and therefore the lower the relative economic benefit from a more accurate ERP forecasting model. In order to evaluate the marginal effect of the risk aversion parameter for the different ERP forecasting models, we reestimate the baseline scenario, for different levels of risk aversion: 1 (more aggressive), 3, 5, and 10 (less aggressive); the other two parameters, that is, the minimum and maximum weights, remain fixed to their default values, $w_{min} = -0.5$ and $w_{max} = 1$.

Fig. 6 presents for all ERP forecasting models (OLS, CT, and PTV) the average equity weight and CER for the different levels of risk aversion. As expected, the higher the level of risk aversion, the lower the average equity weight. As a result, any potential benefit from a more accurate ERP forecasting model shrinks progressively. In the extreme case when relative risk aversion is equal to ten, the various ERP forecasting models have no material difference in their economic impact. Put differently, for a relatively conservative investor, the constrained models do not add any economic benefit above and beyond what is attainable with a simple unconstrained OLS model, as it is justified by the convergence of the graphs, especially for CER, in Fig. 6.

4.2. The effect of minimum weight and short selling

The minimum weight restriction can turn out to be rather critical for the performance of the ERP forecasting models. By construction, the constrained CT and PTV models generate non-negative ERP forecasts. As a result, the respective equity weight is always non-negative and therefore a short selling constraint becomes redundant. In fact, a short selling constraint applied to the unconstrained OLS model is equivalent to using ERP forecasts from the CT model.

In order to evaluate the marginal effect of the minimum weight restriction for the different ERP forecasting models, we reestimate the baseline scenario, for different levels of the minimum weight: -0.5 (allow short selling of up to 50% of the total wealth), -0.3 (allow short selling of up to 30% of the total wealth), 0 (allow no short selling), 0.5 (require minimum equity investment of at least 50% of the total wealth); the other two parameters, that is, the relative risk aversion and the maximum weight, remain fixed to their default values, $\gamma = 3$ and $w_{max} = 1$.

The tighter the minimum weight constraint is, the more likely it is that the constraint becomes binding more often. Panel A of Fig. 7 presents for all ERP forecasting models (OLS, CT, and PTV), the average amount of time (across 804 months of the out-of-sample period), that the minimum weight constraint becomes binding for the different levels of minimum weight. Panel B of Fig. 7 presents the average equity weight and the CER for all forecasting models and different levels of permissible minimum weight.

Trivially, as shown in Fig. 7, the non-positive minimum weight constraints have no effect on the constrained ERP forecasting models, because these models are specifically designed in order to generate non-negative ERP forecasts and

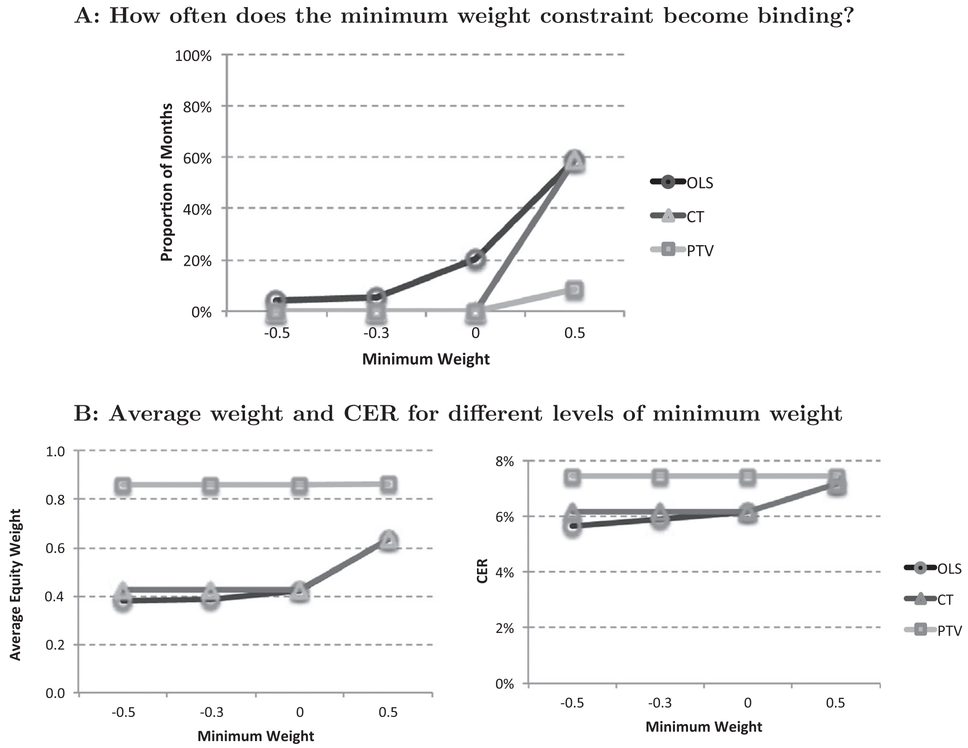

Fig. 7. The effect of minimum weight constraint. The figure presents in Panel A the proportion of months that different levels of the minimum weight constraint become binding, when using out-of-sample ERP forecasts that are generated by the dividend-price ratio based on the unconstrained OLS model, the constrained CT model, and the constrained PTV model. Panel B presents the marginal impact on the average equity weight (left diagram) and on the CER (right diagram) that is implied for different levels of the minimum weight constraint for the various forecasting models. For both panels, the relative risk aversion and the maximum weight are fixed to $\gamma = 3$ and $w_{max} = 1$ respectively. The out-of-sample period is from January 1947 to December 2013.

therefore equity weights. When the minimum weight is positive, which corresponds to a minimum required investment for the investor, the constraint starts becoming binding even for the constrained models, with the proportion of time that this happens being expectedly larger for the CT model than for the PTV model. In fact, for the assumed level of risk aversion ($\gamma = 3$), even a minimum weight constraint of 0.5 rarely becomes binding for the PTV model. This explains the straight lines in Panel B of Fig. 7 for the PTV model.

Overall, the main finding is that for an investor who faces short-sale constraints or a minimum required equity investment, the constrained models can hardly generate a significant economic benefit above and beyond what is attainable with a simple unconstrained OLS model. This can be justified by the convergence of the graphs, especially for CER, in Panel B of Fig. 7.

4.3. *The effect of maximum weight and leverage*

The maximum weight restriction is equivalent to allowing (or otherwise) the use of leverage. When $w_{max} > 1$, the investor is allowed to employ leverage by borrowing at the risk-free rate. In practice, a risk-averse investor might be leverage constrained either due to an investment mandate or because of a broad dislike of leverage and the risks that are associated with it (Gârleanu and Pedersen, 2011; Asness et al., 2012).

In order to evaluate the marginal effect of the maximum weight restriction for the different ERP forecasting models, we reestimate the baseline scenario for different levels of the maximum weight: 0.5 (constrain the portfolio against an all-equity possibility), 1 (allow no leverage), 1.3 (allow leverage of 30% of the total wealth), 1.5 (allow leverage of 50% of the total wealth); the other two parameters, that is, the relative risk aversion and the minimum weight, remain fixed to their default values, $\gamma = 3$ and $w_{min} = -0.5$.

By definition, as the constraint becomes tighter, it becomes more binding. Fig. 8 follows the structure of Fig. 7 and presents in Panel A the average amount of time that the maximum weight constraint becomes binding for different levels of maximum weight and for all ERP forecasting models (OLS, CT, and PTV). Panel B presents the average equity weight and the CER for all forecasting models and different levels of permissible maximum weight.

Panel A of Fig. 8 shows that, contrary to the marginal effect of the minimum weight constraint, the maximum weight constraint becomes more binding for the PTV model, as this generates on average larger ERP forecasts (see the ERP

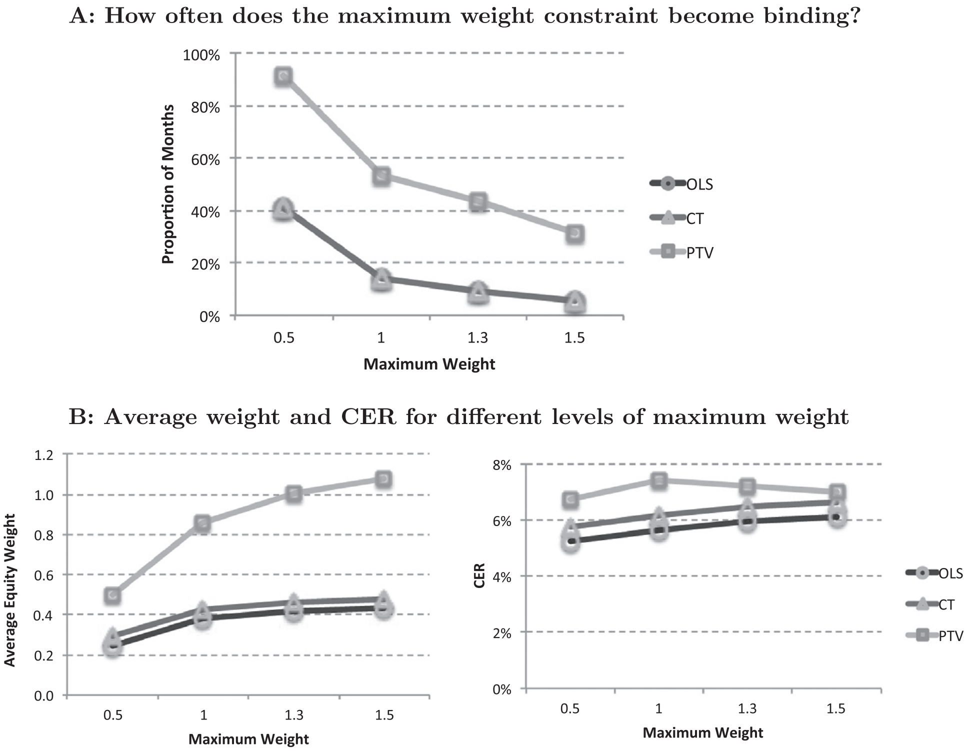

Fig. 8. The effect of maximum weight constraint. The figure presents in Panel A the proportion of months that different levels of the maximum weight constraint become binding, when using out-of-sample ERP forecasts that are generated by the dividend-price ratio based on the unconstrained OLS model, the constrained CT model, and the constrained PTV model. Panel B presents the marginal impact on the average equity weight (left diagram) and on the CER (right diagram) that is implied for different levels of the maximum weight constraint for the various forecasting models. For both panels, the relative risk aversion and the minimum weight are fixed for all the scenarios to $\gamma = 3$ and $w_{min} = -0.5$ respectively. The out-of-sample period is from January 1947 to December 2013.

histograms of Fig. 2). Trivially, the unconstrained OLS model and the constrained CT model are equally affected by the maximum weight constraint, because their ERP forecasts are identical when positive. The left diagram of Panel B of Fig. 8 shows that the average equity allocation increases significantly, as the maximum weight constraint becomes looser (i.e., as w_{max} increases). Interestingly, the PTV model in particular, leads on average to leveraged equity allocations (i.e., average weight exceeds 1.0), when the maximum weight constraint is 1.3 or above.

The right diagram of Panel B in Fig. 8 shows that the CER typically increases when the maximum weight constraint increases. However, for the PTV model, higher levels of leverage do not necessarily lead to larger CER. The PTV model no longer offers higher levels of CER when leverage is allowed (i.e., when $w_{max} > 1.0$); one can notice the kink in the plot.

4.3.1. *A closer look at the dependence of the PTV model on leverage*

We look at two issues related to the PTV model and the potential implications of associated leverage. First, we look closely at the relation between the times that the maximum weight constraint becomes more binding and the underlying market regime. Second, we explore the lack of additional economic benefit in terms of CER, when leverage is allowed, as deduced from the kink in the PTV plot in Panel B of Fig. 8.

To comprehend the relation between market regime and the frequency that a constraint becomes binding, we go back to Eq. (14) and the definition of the equity weight in our asset allocation framework. The equity weight is proportional to the actual ERP forecast of each model and inversely proportional to the risk aversion of the investor and the historical equity market variance; the latter two parameters are obviously common across all forecasting models that we study. Thus, given a certain level of risk aversion, the maximum weight constraint is more likely to become binding during periods of low realized equity market variance and/or periods of large ERP forecasts.

Broadly speaking, the equity market variance increases during market drawdowns and recessionary periods. As long as the various ERP forecasting models do not generate large enough forecasts to compensate for the higher levels of returns variance in stress environments, the actual equity weight is unlikely to reach such high levels for the maximum weight constraint to become binding. In unreported results, we find that there is no systematic pattern with regards to the periods when the maximum weight constraint becomes binding for the PTV model.

To validate the above argument, we note that if the maximum weight constraint becomes binding during an up market, it

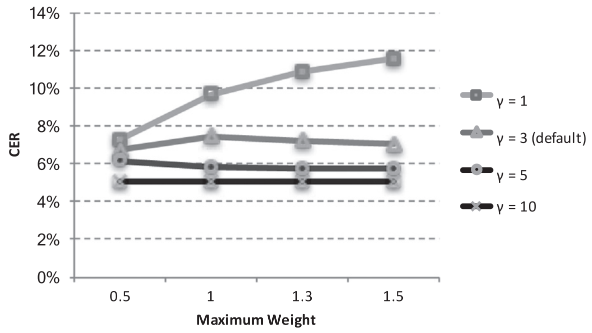

Fig. 9. Economic profitability and the maximum weight constraint for the PTV model. The figure presents the CER that is implied for different levels of the maximum weight constraint for different levels of relative risk aversion (γ = 1, 3, 5, 10), when using out-of-sample ERP forecasts that are generated by the dividend-price ratio and the constrained PTV model. The minimum weight is fixed for all the scenarios to w_{min} = – 0.5. The out-of-sample period is from January 1947 to December 2013.

is actually hurting the upside of the asset allocation framework, whereas if the constraint becomes binding during a down market, it is actually protecting the investor from employing leverage in a stress market. Put differently, the maximum weight constraint acts as an upside cap in good market environments and as a stop loss rule during periods of stress, especially for constrained models like the PTV model.

Table 5 reports CER gains for the PTV model (versus the OLS model) of −6.70% for recessions (Panel D) and 3.34% for expansions (Panel E), under our baseline set of parameters ($\gamma = 3$, $w_{min} = -0.5$, $w_{max} = 1$). If we allow for leverage, by setting $w_{max} = 1.3$, these estimates become −11.66% and 3.60% respectively; for $w_{max} = 1.5$, we get −14.05% and 3.63% respectively.[19] Thus, we document a symmetrical behaviour; the looser the maximum weight constraint becomes, the greater the benefit during good markets (stronger outperformance of the constrained PTV model versus the unconstrained OLS model) and the greater the penalty during bad markets (stronger underperformance of the constrained model).

Next, we investigate why higher levels of leverage do not necessarily lead to larger CER for the PTV model, as illustrated in Panel B of Fig. 8. Broadly speaking, the use of leverage trivially increases the impact of positive equity returns, but it also magnifies the volatility of returns. This can, as already shown, have an adverse effect in the economic profitability of the strategy, when assessed in terms of the CER. By definition – see Eq. (16) – the CER is equal to the average return of the strategy, penalized for the associated variance. The sensitivity on portfolio variance is controlled by the relative risk aversion γ of the investor. Along these lines, a more risk-averse investor (higher γ) would require a larger increase in average returns to compensate for the higher level of volatility due to leverage.

Returning to Panel B in Fig. 8 and the PTV model, when $w_{max} = 1$, the strategy delivers an average annualized return of 9.93% with a volatility of 12.86%, hence leading to a CER of 7.45% for an investor with $\gamma = 3$. Allowing for leverage, and taking as an example the case when $w_{max} = 1.3$, increases the average return to 10.74%, but the increase in the volatility (to 15.36%) goes beyond what the investor with $\gamma = 3$ would allow for the CER to remain at least unchanged; the CER in that case is 7.21%, hence causing the kink in the PTV plot. For completeness, the break-even increase in volatility for the CER to remain at the levels of 7.45% is 14.81%. Alternatively, the CER would be the same if the investor became slightly more risk loving and exhibited a risk aversion of 2.8.

Another way to present the same result is by looking at the relative return of the PTV model, against the OLS model (the result remains the same if the comparison is done against the CT model). The PTV model always outperforms the OLS model in terms of average return, and most importantly, the differential return increases as more leverage is allowed. When $w_{max} = 1$, the return differential between the PTV model and the OLS model is 3.05% and it increases to 3.28% and 3.48% when w_{max} is 1.3 and 1.5 respectively. However, these increases in the relative performance gain come at significantly larger levels of volatility for the PTV compared to the OLS model. This leads to progressively smaller economic gains. The difference in the CER estimates, which is the CER gain as defined in Eq. (17), is 1.80%, 1.24%, and 0.90%. This can be seen by the progressively smaller distance between the PTV and OLS plots in the right diagram of Panel B in Fig. 8.

In sum, leverage has a dual impact: (1) it increases the outperformance of the more constrained models in terms of average returns, but (2) it can lead to smaller relative economic benefits for more risk-averse investors. In Fig. 9, we show the results of an analysis of the right diagram of Panel B in Fig. 8 for different levels of risk aversion; the default case is $\gamma = 3$. The results demonstrate the dynamics between leverage, CER, and risk aversion.

For a largely risk-averse investor with $\gamma = 10$, the allocation to the equity market never exceeds 0.5 in our sample period and therefore the various levels of w_{max} that we use in our analysis are irrelevant; the CER is constant, as if there was no maximum weight constraint. Conversely, for a risk-loving investor with $\gamma = 1$, the more leverage is allowed the better in

[19] In unreported results, we find that the effects are of similar nature and of larger magnitude if instead of recessionary and expansionary periods, we focus on the up and down markets.

terms of economic profitability; the CER is monotonically increasing. Between these limiting cases, we have two cases of investors who experience the kink in their CER plot at different levels of the w_{max}. The investor with $\gamma = 3$ as already discussed, dislikes leverage; the kink is at $w_{max} = 1$. As the investor becomes more risk averse, the level of w_{max} at which the kink is witnessed falls. For the investor with $\gamma = 5$, this is at $w_{max} = 0.5$; the investor is better off even when constrained against a 100% allocation in the equity market.

Overall, taken together, the main findings of this sensitivity analysis show that the relative profitability of a dynamic strategy that allocates between the equity market and a risk-free asset using econometrically superior ERP forecasts shrinks significantly if the investor is highly risk-averse and/or heavily constrained.

5. Conclusion

Researchers have long been occupied with suggesting models and macroeconomic variables to forecast excess market returns more accurately. We ask an important question: Does econometric predictability across the business cycle imply predictability at all times? Our aim has been to highlight the importance of conditional as opposed to unconditional performance evaluation of ERP forecasting models. Our findings raise concerns about the real value-add of recent academic advances in ERP forecasting models when these are evaluated across different market regimes and realistic investment constraints are taken into account.

We use the Campbell and Thompson (2008) and Pettenuzzo et al. (2014) constrained models, as a good example of models that provide more accurate ERP forecasts over a long sample period and across the business cycle. Our focus is to investigate the benefits of these models across various market regimes for a risk-averse asset allocator who dynamically allocates capital in the equity market.

We show that during periods of high volatility and market downturns, the econometrically superior constrained Campbell and Thompson (2008) and Pettenuzzo et al. (2014) forecasting models lead to large relative economic losses, compared to unconstrained models. Conversely, we find that than any relative economic benefit that is mainly occurring during up markets, can be significantly reduced, if the investor is highly risk-averse and faces leverage constraints.

Overall, we illustrate that greater unconditional econometric predictability does not necessarily lead to superior predictability at all times, especially when financial intermediation idiosyncrasies (aversion to relative losses in down markets so to avoid outflows; investment constraints imposed by regulation or mandate) are taken into account. New ERP forecasting frameworks should aim to accommodate these features. The regime-switching model of Zhu and Zhu (2013) and the state-dependent model of Huang et al. (2017) are, to our knowledge, the first explicit attempts in this direction. An alternative suggestion would favour an optimal combination of unconstrained and constrained forecasting models in a dynamic framework, driven by market regime indicators; Ang and Bekaert (2002, 2004), Guidolin and Timmermann (2008), Tu (2010), and Zhu and Zhu (2013) can provide guidance in that respect. We leave this for future research.

References

Ang, A., Bekaert, G., 2002. International asset allocation with regime shifts. Rev. Financ. Stud. 15, 1137–1187.
Ang, A., Bekaert, G., 2004. How regimes affect asset allocation. Financ. Anal. J. 60, 86–99.
Ang, A., Bekaert, G., 2007. Stock return predictability: is it there? Rev. Financ. Stud. 20, 651–707.
Asness, C.S., Frazzini, A., Pedersen, L.H., 2012. Leverage aversion and risk parity. Financ. Anal. J. 68, 47–59.
Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Rev. Financ. Stud. 22, 4463–4492.
Boudoukh, J., Richardson, M., Whitelaw, R.F., 2008. The myth of long-horizon predictability. Rev. Financ. Stud. 21, 1577–1605.
Brown, S., 2008. Elusive return predictability: discussion. Int. J. Forecast. 24, 19–21.
Campbell, J.Y., Shiller, R.J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Rev. Financ. Stud. 1, 195–228.
Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Rev. Financ. Stud. 21, 1509–1531.
Campbell, J.Y., Yogo, M., 2006. Efficient tests of stock return predictability. J. Financ. Econ. 81, 27–60.
Cashman, G.D., Deli, D.N., Nardari, F., Villupuram, S., 2012. Investors do respond to poor mutual fund performance: evidence from inflows and outflows. Financ. Rev. 47, 719–739.
Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. J. Econ. 138, 291–311.
Cochrane, J.H., 2011. Presidential address: discount rates. J. Financ. 66, 1047–1108.
Cooper, I., Priestley, R., 2009. Time-varying risk premiums and the output gap. Rev. Financ. Stud. 22, 2801–2833.
Dangl, T., Halling, M., 2012. Predictive regressions with time-varying coefficients. J. Financ. Econ. 106, 157–181.
Fama, E.F., French, K.R., 1988. Dividend yields and expected stock returns. J. Financ. Econ. 22, 3–25.
Ferreira, M.A., Santa-Clara, P., 2011. Forecasting stock market returns: the sum of the parts is more than the whole. J. Financ. Econ. 100, 514–537.
Fung, W., Hsieh, D.A., Naik, N.Y., Ramadorai, T., 2008. Hedge funds: performance, risk, and capital formation. J. Financ. 63, 1777–1803.
Gârleanu, N., Pedersen, L.H., 2011. Margin-based asset pricing and deviations from the law of one price. Rev. Financ. Stud. 24, 1980–2022.
Goyal, A., Welch, I., 2003. Predicting the equity premium with dividend ratios. Manag. Sci. 49, 639–654.
Goyal, A., Welch, I., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21, 1455–1508.
Guidolin, M., Timmermann, A., 2008. International asset allocation under regime switching, skew, and kurtosis preferences. Rev. Financ. Stud. 21, 889–935.
Hansen, P.R., Timmermann, A., 2012. Choice of sample split in out-of-sample forecast evaluation. Working Paper.
Henkel, S.J., Martin, J.S., Nardari, F., 2011. Time-varying short-horizon predictability. J. Financ. Econ. 99, 560–580.
Henriksson, R.D., Merton, R.C., 1981. On market timing and investment performance. ii. statistical procedures for evaluating forecasting skills. J. Bus. 54, 513–533.
Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: a powerful predictor of stock returns. Rev. Financ. Stud. 28, 791–837.

Huang, D., Jiang, F., Tu, J., Zhou, G., 2017. Forecasting stock returns in good and bad times: the role of market states. SSRN eLibrary.
Huang, J., Wei, K.D., Yan, H., 2007. Participation costs and the sensitivity of fund flows to past performance. J. Financ. 62, 1273–1311.
Ivković, Z., Weisbenner, S., 2009. Individual investor mutual fund flows. J. Financ. Econ. 92, 223–237.
Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: why imposing the wrong constraints helps. J. Financ. 58, 1651–1684.
Kandel, S., Stambaugh, R.F., 1996. On the predictability of stock returns: an asset-allocation perspective. J. Financ. 51, 385–424.
Leitch, G., Tanner, J.E., 1991. Economic forecast evaluation: profits versus the conventional error measures. Am. Econ. Rev. 81, 580–590.
Li, J., Yu, J., 2012. Investor attention, psychological anchors, and stock return predictability. J. Financ. Econ. 104, 401–419.
Li, Y., Ng, D.T., Swaminathan, B., 2013. Predicting market returns using aggregate implied cost of capital. J. Financ. Econ. 110, 419–436.
Lo, A.W., 2004. The adaptive markets hypothesis. J. Portf. Manag. 30, 15–29.
Møller, S.V., Rangvid, J., 2015. End-of-the-year economic growth and time-varying expected returns. J. Financ. Econ. 115, 136–154.
Neely, C.J., Rapach, D.E., Tu, J., Zhou, G., 2014. Forecasting the equity risk premium: the role of technical indicators. Manag. Sci. 60, 1772–1791.
O'Neal, E.S., 2004. Purchase and redemption patterns of US equity mutual funds. Financ. Manag. 33, 63–90.
Pesaran, M.H., Timmermann, A., 1992. A simple nonparametric test of predictive performance. J. Bus. Econ. Stat. 10, 461–465.
Petajisto, A., 2013. Active share and mutual fund performance. Financ. Anal. J. 69, 73–93.
Pettenuzzo, D., Timmermann, A., Valkanov, R., 2014. Forecasting stock returns under economic constraints. J. Financ. Econ. 114, 517–553.
Rapach, D.E., Ringgenberg, M.C., Zhou, G., 2016. Short interest and aggregate stock returns. J. Financ. Econ. 121, 46–65.
Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Rev. Financ. Stud. 23, 821–862.
Sirri, E.R., Tufano, P., 1998. Costly search and mutual fund flows. J. Financ. 53, 1589–1622.
Timmermann, A., 2008. Elusive return predictability. Int. J. Forecast. 24, 1–18.
Tu, J., 2010. Is regime switching in stock returns important in portfolio decisions? Manag. Sci. 56, 1198–1215.
Zhu, X., Zhu, J., 2013. Predicting stock returns: a regime-switching combination approach and economic links. J. Bank. Financ. 37, 4120–4133.