Forecasting the Equity Risk Premium: the Ups and the Downs*

NICK BALTAS[†] AND DIMITRIOS KARYAMPAS[‡]

August 8, 2016

ABSTRACT

Asset allocation is critically dependent on the ability to forecast the equity risk premium (ERP) out-of-sample. Constrained forecasting models have been recently introduced in order to improve the poor out-of-sample forecasting ability of macroeconomic variables like the dividend-price ratio. This paper critically investigates the nature of these constraints and their implications for dynamic asset allocation. Consistent with the existing evidence, such models improve the economic benefit for a mean-variance investor over a very long sample period (1947-2013). However, seen from a conditional viewpoint, we show that constrained models generate significant economic relative losses in periods of high volatility and market drawdowns, when it matters the most for asset allocators to retain assets and client base. Additionally, we find that risk-averse investors that face investment constraints –either by mandate or regulation– like short-selling or leverage constraints, can find little benefit in constrained ERP forecasting models, even across the business cycle. Our findings pose a significant challenge on the practical application of constrained ERP forecasting models and call for new model designs that actively incorporate some form of regime dependency.

JEL CLASSIFICATION CODES: C53, C58, G11, G14, G17

KEYWORDS: equity risk premium predictability, out-of-sample forecasting, economic constraints, predictive regression, asset allocation, business cycles, trading strategies

---

*Comments by Mathijs Cosemans, Simon Scheidegger and Guofu Zhou and participants at the 2015 Computational and Financial Econometrics (CFE) conference are gratefully acknowledged. Further comments are warmly welcomed, including references to related papers that have been inadvertently overlooked. The views expressed in this article are those of the authors only and no other representation should be attributed. All errors are our own.

[†]Imperial College Business School, South Kensington Campus, SW7 2AZ, London, UK; e-mail: n.baltas@imperial.ac.uk.

[‡]ICMA centre, University of Reading, RG6 6BA, Reading, United Kingdom; e-mail: karyampas@gmail.com.

1. Introduction

Asset allocation decisions, whether strategic or tactical, are closely related to forecasts of the equity risk premium (ERP, henceforth). This renders the ERP forecasting model of critical practical importance for portfolio managers and asset allocators. Importantly enough, in making their investment decisions, these investment professionals typically face a number of constraints, like short-selling or leverage constraints, which might significantly shrink the value add of a sophisticated ERP forecasting model. The objective of this paper is twofold. First, we critically evaluate the out-of-sample forecasting performance of unconstrained and –recently introduced– constrained ERP forecasting models for an investor with mean-variance preferences. Second, we carefully investigate the practical implications of realistic investment constraints when making tactical asset allocations based on ERP forecasts of the various models under consideration.

The academic literature in financial economics has long been occupied with the challenging empirical task of forecasting equity market returns. Even though a long list of macroeconomic variables have been shown to forecast equity market returns on an in-sample basis using a univariate linear regression framework[1], these variables fail to generate more accurate ERP forecasts than the simple ERP historical average on an out-of-sample basis, as first highlighted by Goyal and Welch (2008). Evidently, out-of-sample –as opposed to in-sample– predictability is of actual economic benefit in dynamic portfolio decision making and for this reason we solely focus on out-of-sample predictability.[2]

In order to improve the out-of-sample forecasting ability of the macroeconomic variables,

[1]Without the list being exhaustive, the most prominent ERP macroeconomic predictors have historically been the treasury bill rate (Fama and Schwert 1977, Breen, Glosten and Jagannathan 1989, Ang and Bekaert 2007), the dividend yield (Fama and French 1988, Campbell and Yogo 2006, Ang and Bekaert 2007), the dividend-price ratio (Campbell and Shiller 1988), term spread (Keim and Stambaugh 1986, Campbell 1987, Fama and French 1988), the equity market volatility (French, Schwert and Stambaugh 1987), the book-to-market ratio (Kothari and Shanken 1997, Pontiff and Schall 1998), the inflation (Fama and Schwert 1977, Campbell and Vuolteenaho 2004), the default spread (Fama 1986, Keim and Stambaugh 1986), corporate issuing activity (Baker and Wurgler 2000), consumption-surplus ratio (Campbell and Cochrane 1999), consumption-wealth ratio (Lettau and Ludvigson 2001). Recent additions to this list, especially after Goyal and Welch (2008) highlight the importance of out-of-sample forecasting, include the output gap (Cooper and Priestley 2009), the variance risk premium (Bollerslev, Tauchen and Zhou 2009), the nearness to the 52-week high and to the historical high (Li and Yu 2012), the aggregate implied cost of capital (Li, Ng and Swaminathan 2013), price and volume based technical indicators (Neely, Rapach, Tu and Zhou 2014), the end-of-year economic growth (Møller and Rangvid 2015), an investor sentiment indicator (Huang, Jiang, Tu and Zhou 2015), a mean-reversion indicator (Huang, Jiang, Tu and Zhou 2016) and short interest (Rapach, Ringgenberg and Zhou 2016).

[2]The assumption of a constant relationship between ERP and a predictor variable could support the generalisation of a significant in-sample relationship into an out-of-sample one, but recent empirical evidence (e.g. Cochrane 2011) has shown that such a relationship is time-varying and state-dependent (Huang et al. 2016).

several methodological refinements of the linear forecasting model have been recently introduced. First, Campbell and Thompson (2008) suggest the introduction of simple constraints in the linear ordinary least squares (OLS, henceforth) forecasting model. One of the most comprehensive constraints is to truncate the out-of-sample 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, Timmermann and Valkanov (2014) introduce a Bayesian forecasting framework, under which conditional out-of-sample ERP forecasts (or the market Sharpe ratio) are constrained to a positive territory (the PTV model, henceforth). Both model refinements manage to significantly improve the out-of-sample forecasting ability of conventional macroeconomic ERP predictors, at least unconditionally. However, the authors of both papers seem to agree that the improvement has been mainly concentrated during earlier years of their respective sample period, and not over the most recent decades, which have been characterised by a number of substantial downturns such as the burst of the dot-com bubble, the credit crisis and the Eurozone crisis.

The first question that we attempt to answer in this paper is whether the introduction of such constrained models does actually improve the performance of dynamic asset allocation. Our hypothesis is that the constrained models can only empirically improve the accuracy of the unconstrained OLS model on an unconditional basis, simply because the average ERP realisation over the business cycle has historically been positive. However, on a conditional basis, during periods of high volatility and negative ERP realisations, such constrained models that generate non-negative (and less volatile) ERP forecasts, are deemed to underperform.

Using a long data sample that spans the period between January 1927 to December 2013 and the dividend-price ratio as the predictor of next month's S&P 500 excess returns, we first confirm the findings of Campbell and Thompson (2008) and Pettenuzzo et al. (2014) in that the constrained models do improve the out-of-sample forecasting 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. In particular, 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, but 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.

In order to assess the practical importance of our findings we conduct an asset allocation analysis, under which a risk-averse investor dynamically allocates between a risky asset (equity market) and a risk-free asset, using out-of-sample ERP forecasts of the various models under consideration. We find that the constrained models generate positive certainty equivalent return 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 for investment professionals to deliver higher relative returns so to retain assets and their client base intact.

Given the above findings, our second research question looks at the effect of constraints from a different perspective. Investors typically face a number of investment constraints, either by mandate or regulation, which put hard threshold on minimum and maximum allocation across risky assets. For instance, 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. Our second objective is to investigate the value add of the various ERP forecasting models under consideration, when the mean-variance portfolio manager faces realistic investment constraints. Our hypothesis is that the tighter the investment constraints that a manages faces, the weaker and less significant the impact of a constrained ERP forecasting model.

Our empirical analysis shows that, on the one hand, a short-sale constraint is –trivially– equivalent to a non-negative ERP forecast and therefore becomes, by construction, redundant for the constrained CT and the PTV models; put differently, the constrained models have little (or even nothing in the case of the CT model) to add, when the manager is by mandate constrained against shorting. On the other hand, a leverage constraint becomes binding for the constrained models, which typically generate larger ERP forecasts than the unconstrained model, hence diminishing their value add in periods of high ERP realisations. All in all, we find that the tighter the investment constraints become and the more risk averse the investor is, 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 for a constrained risk-averse investor, than for an investor who can take short positions or most importantly employ leverage.

Taken together, our findings show that constraining the ERP forecasting model does not seem to be of significant added value for constrained asset allocators on a dynamic basis. Consequently, these findings have important implications for the design of new ERP forecasting models as well as the identification of new predictor variables, and they clearly highlight the requirement for accommodating some form of regime dependency. Failing to forecast market downturns especially during volatile periods, exactly when it is needed the most, can be detrimental for actual portfolio decisions as already explained. Huang et al. (2016) are possibly the first –to our knowledge– to explicitly highlight the need for forecasting models that can accurately forecast future ERP across both good and bad times. We discuss these issues in more detail at a later stage in the paper.

The paper is organised 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

This section provides an outline of the baseline linear forecasting model of market excess returns as well as of its recent constrained variants introduced first by Campbell and Thompson (2008) and subsequently by Pettenuzzo et al. (2014). Additionally, the section provides an overview of our dataset.

For the main core of our analysis, we focus on out-of-sample ERP forecasting at the monthly horizon, as predictability across longer horizons can potentially be an artefact of highly persistent predictor variables (Boudoukh, Richardson and Whitelaw 2008, Cochrane 2011). Longer-term ERP forecasting can be of practical importance for investors that have longer investment horizons, like pension funds. For this reason and for additional robustness, we have also performed analysis for quarterly and annual horizons. The results are in line with the findings presented in the paper and are available upon request.

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 out-of-sample 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:

$$r_{\tau+1} = \alpha + \beta \cdot x_\tau + \varepsilon_{\tau+1}, \quad \tau = 1, \cdots, t-1 \tag{2}$$

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:

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

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. CT show that the tructated model improves significantly the out-of-sample forecasting performance of a large list of predictor variables.

Following the same basic economic reasoning of Campbell and Thompson (2008) on the positivity of the forecasted ERP, Pettenuzzo et al. (2014) attempt to improve the methodology and 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 in that they lead to positive in-sample and out-of-sample forecasted excess returns:

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

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; we refer the reader to the Appendix (Sections A.1 & A.2) of Pettenuzzo et al. (2014). The way that the PTV model differs from the CT model is that it performs a proper constrained optimisation in the fitting of the model instead of enforcing a post-fitting truncation.

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 entire academic literature on ERP forecasting has used the historical average ERP as the benchmark model. 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 therefore specifically 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 later in the paper, 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 make use of 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 empirical analysis is conducted over a long sample period, 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&P500 index, as maintained by the Center of Research for Security Prices (CRSP). Without loss of generality, we present our findings using the dividend-price ratio (d/p) of the index as the predictor variable; the historical values of d/p is shown in Panel A of Figure 1. Monthly data for the total returns of the S&P500 as well as for d/p are collected from the website of Amit Goyal[3].

Our analysis does not constitute a comparison study between different predictor variables, but instead a comparison between the mechanics of different forecasting methodologies. For that purpose, the choice of the predictor that we use (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).[4]

[Figure 1 about here]

Over the entire sample period the average annualised ERP has been 5.90% with a volatility of 19.15% (Sharpe ratio of 0.31) and a negative skewness of -0.42.

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

[4] These results are omitted for reasons of space, but are available upon request from the authors.

3. Out-of-sample Forecasting

The aim of this section is to empirically evaluate the forecasting ability between constrained and unconstrained ERP forecasting models. In order to remain consistent with the existing literature, 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 from January 1947 onwards.

To start with, Panel B of Figure 1 presents the monthly out-of-sample ERP forecasts of the three models, the unconstrained OLS regression model, the CT truncation model and the PTV Bayesian model[5]. 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 binding; this is roughly during the period 1992-2005. Contrary to OLS and CT forecasts, the PTV forecasts remain strictly positive across the entire sample period and exhibit relatively smaller time-series variation.

In order to summarise the ERP forecasting behaviour of the unconstrained OLS model and the constrained CT and PTV models, Figure 2 presents the histograms of the 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 ERP forecast and at the same time reduces the standard deviation around this average level. This appears to come at odds with Huang et al. (2015), who argue that more volatile ERP forecasts are more likely to track the largely volatile ERP realisations more closely.

[Figure 2 about here]

It seems obvious, at least numerically, that the less constrained models are more likely to generate more accurate pointwise ERP forecasts in periods when the realisation of the market excess return happens to be negative. On the contrary, the more constrained models would provide more accurate pointwise ERP forecasts when the realisation of the market excess return happens to be positive in general. It must be stressed that the discussion in this paragraph has remained at a philosophical and a qualitative level, yet it serves as our motivation for the

[5]The PTV forecasts can be directly compared with Figure 9 of Pettenuzzo et al. (2014).

analysis that follows in the next pages.[6]

Having acknowledged the potential pitfalls of a heavily constrained model, we can hypothesise that the statistical forecasting outperformance of the constrained models, as presented in the results of Campbell and Thompson (2008) and Pettenuzzo et al. (2014), can be potentially driven by the fact that the realised 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 Figure 2 appears much closer to the constrained models than the unconstrained ones.

Arguably, the level of closeness of the average ERP forecast to the average historical ERP realisation is not enough to deem one model superior to another, so we next proceed with presenting statistical evidence.

3.1. Forecasting accuracy against the historical average

Our main objective is to compare constrained ERP forecasting models with the unconstrained OLS model. Before presenting these results and in order to be consistent with the academic literature on ERP forecasting, we first briefly present how all these forecasting models compare with the historical ERP average forecast, which constitutes the standard forecasting benchmark.

In that respect and following Goyal and Welch (2008) and Campbell and Thompson (2008), we compare the forecasting errors of all models, 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 Figure 3 presents the Difference in the Cumulative Sum of Squared Errors (*DCSSE*) between the OLS, CT and PTV[7] models and the

[6]It is critical to note that the described dependencies between the ERP forecasting models and the ERP realisations aim to illustrate the average expectation in terms of accuracy. This, however, cannot line out a contradicting scenario, under which an unconstrained OLS model turns out to be very poor in forecasting market excess returns, if it forecasts large and positive ERP when this happens to be very negative and vive versa. Along these lines, what we merely aim to illustrate at this point is that a model that forecasts negative ERP can, under reasonable conditions and over the business cycle, be more precise than a constrained model during negative-return months.

[7]The results for the PTV model can be directly compared with Figure 13 of Pettenuzzo et al. (2014).

historical average:

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

where $Model = \{OLS, CT, PTV\}$. An upward movement of the $DCSSE_{HA}$ indicates that the model is more accurate in forecasting the ERP compared to the historical average. Conversely, when the $DCSSE_{HA}$ moves downwards, the historical average is more accurate. Trivially, if the $DCSSE_{HA}$ remains flat, the performance of the model of interest is similar to that of the historical average.[8]

[Figure 3 about here]

The evidence shows that the constrained models appear to improve the forecasting accuracy of the unconstrained OLS model over the entire sample period, even though there exist periods during which the latter generates significantly lower forecast errors. Regarding the strictness level of the ERP constraint that is employed, the CT model appears to improve the performance of the OLS model following periods when the hard zero-bound constraint becomes binding (as identified in Figure 2) in line with Campbell and Thompson (2008). The PTV model, which employs an even 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. In particular the $DCSSE_{HA}$ path increases strongly up until around 1956, then increases at a much lower pace up until 1995, before falling up until around 2000 and remaining flat ever since.

Put simply, the constrained models appear to improve the forecasting power of the unconstrained model, in line with Campbell and Thompson (2008), Pettenuzzo et al. (2014), but this improvement tends to be achieved during earlier periods of the sample. The most recent decades constitute a significant challenge for the performance of the constrained models. We next proceed with a more direct comparison between the constrained and the unconstrained models, which constitutes the core of our analysis.

[8]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).

3.2. Forecasting accuracy against the unconstrained OLS model

Setting aside the historical average, we turn our attention to the main part of our analysis, which is to evaluate the benefits and pitfalls of a constrained forecasting model against the baseline unconstrained OLS model.

For that purpose, Table I (column "All") presents for both predictors, the root mean square forecast error (RMSFE) of the three 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 than the unconstrained OLS model across the entire sample.[9] We find that both constrained models improve the forecasting ability of the unconstrained model. The $R^2_{OOS,OLS}$ is 0.18% for CT model and 0.57% for PTV, both statistically significant.

[Table I about here]

In order to visualise these findings, Panel B of Figure 3 presents the difference in the cumulative sum of squared forecast errors between the constrained models, CT and PTV, and the forecasting errors of 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 Equation 9, but allows for the direct comparison of the accuracy between the unconstrained OLS model and its constrained variants. When the $DCSSE_{OLS}$ moves upwards over time, this indicates that the constrained model is more accurate in forecasting the ERP compared to the unconstrained one. Conversely, when the $DCSSE_{OLS}$ moves downwards, the unconstrained model is more accurate.

The evidence is completely in line with our hypothesis. The constrained CT and PTV models manage to improve the forecasting performance of the unconstrained OLS model (the respective $DCSSE_{OLS}$ lines end up in the positive territory at the end of the sample period and, as

[9] The statistic 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 to that of the OLS model. The statistic 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 statistic for the CT model would largely increase. As an example, the estimates for the CT model in Table I would increase from 13.20% to 92.90% .

already discussed, the $R^2_{OOS,OLS}$ values are all positive in Table I), but there exist several periods during which the unconstrained model is more accurate.

The constrained models appear to outperform the baseline OLS model during strong market rallies, but significantly underperform during volatile periods of negative ERP realisations. The 36-month rolling Sharpe ratio of the market that is presented in Figure 4 can assist in supporting this argument. At times when the Sharpe ratio falls (due to either low ERP realisations 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.

[Figure 4 about here]

The most recent decades can highlight this prescribed dependence between the performance of the constrained models and the ERP realisations even more clearly. The build-up of the dot-com bubble in the late 1990's with the subsequent collapse up until the last months of 2002, the bull market of the following years, with the subsequent credit crisis in 2008 and the most recent market rally that was temporarily hit by the Eurozone crisis episodes in 2010 and 2011 are patterns that are present in both the rolling Sharpe ratio calculation, as well as in the forecast errors of the constrained models relative to the unconstrained OLS model.

To summarise, our full-sample results confirm the findings of Campbell and Thompson (2008) and Pettenuzzo et al. (2014) in that constrained forecasting models improve the forecasting ability of an unconstrained model. However, the constrained forecasting models appear to significantly benefit during up markets, whereas they seem to hurt during periods with negative ERP realisations. We therefore continue our analysis with a careful evaluation of the forecasting performance of the models conditional on various market regimes.

3.3. Forecasting accuracy and the market regime

The last two columns of Table I present the forecasting performance statistics separately for up and down markets, which are trivially defined by months with positive and negative ERP realisation. Over the entire out-of-sample period, January 1947 to December 2013, we identify 476 months with a positive ERP realisation and 328 months with a negative ERP realisation.

The empirical evidence is overwhelming. During up markets, the constrained forecasting 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 academic 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. Importantly enough, the PTV model generates significantly larger $R^2_{OOS,OLS}$ values than the CT model, which effectively shows that the stricter the constraint, the stronger the outperformance. On the contrary, during periods of negative ERP realisation, the constrained models substantially increase the forecasting error of the unconstrained OLS model, hence resulting in negative and large in magnitude $R^2_{OOS,OLS}$, with the stricter PTV model suffering more than the CT model in complete symmetry to the up market result.

From a statistical perspective, it's important to infer whether during down markets the constrained models significantly underperform the unconstrained OLS model. The Clark and West (2007) MSFE-*adjusted* statistic can be used to test for the null hypothesis $H_0' : R^2_{OOS,OLS} \geq 0$ against the one-sided, lower-tail, alternative hypothesis hypothesis $H_A' : R^2_{OOS,OLS} < 0$; notice that this is the complete opposite statistical test than the one reported in all the Tables of our paper, hence the use of the prime in the notation. In order to avoid confusion, we don't report the statistical significance of the $R^2_{OOS,OLS}$ values during down markets in Table I, 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.[10] In other words, during down markets, the constrained models suffer dramatically when compared to the simple unconstrained OLS model.

Overall, these findings are very sound. The constrained ERP forecasting models significantly outperform the unconstrained model during months of positive ERP realisation and significantly underperform the unconstrained model during months of negative ERP realisation. This state-dependent behaviour is visualised in Figures 5, which extends Panel B of Figure 3 and present the $DCSSE_{OLS}$ calculation separately for months with positive EPR realisation

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

($DCSSE^{+}_{OLS}$) and for months with negative ERP realisation ($DCSSE^{-}_{OLS}$):

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

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

where $I_{r_\tau \geq 0}$ and $I_{r_\tau < 0}$ denote indicator functions for positive and negative ERP realisations.

[Figure 5 about here]

The empirical evidence from Figure 5 is indeed very strong and justifies the magnitude of the $R^2_{OOS,OLS}$ values of Table I. By conditioning the evaluation of the forecasting performance of the models on the sign of the ERP realisation, we confirm our hypothesis in that the unconstrained OLS model outperforms the constrained models in months of negative ERP realisation, whereas the constrained models outperform the unconstrained OLS model in months of positive ERP realisation.

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 realisation, but additionally on the magnitude of the return. In particular, starting from the subset of months with a positive ERP realisation and from the subset of months with a negative EPR realisation, we subsequently divide equally the negative ERP return bucket into negative-large (NL) and negative-small (NS) subsets, and equivalently divide the positive ERP return bucket into positive-large (PL) and positive-small (PS) subsets. Table II and presents the same forecasting performance statistics as in Table I across all four regimes (NL, NS, PS and PL).

[Table II about here]

In line with our findings so far, Table II shows that the constrained models strongly outperform the unconstrained OLS model over months with positive ERP realisations, but also strongly underperform the unconstrained OLS model over months with negative ERP realisations (NL and NS).

The constrained models appear to perform best, relative to the unconstrained model, in the PS regime, i.e. when the ERP is positive and small, in the range 0% to 2.9%. The $R^2_{OOS,OLS}$

reaches extremely large and statistically significant values, all at 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 realisations range from 2.9% up 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 very sound, 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 struggle 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 negative and has small magnitude, in the range -2.4% to 0%. 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 realisations. In fact, the fact that the constrained models suffer significantly more, in relative terms to the unconstrained OLS model, during the NS regime can turn out to be detrimental for an investor who allocates between the equity markets and a risk-free asset, as she could potentially suffer large utility losses if the market realises even a small negative return. We leave the discussion of any asset allocation implications of the various forecasting models for Section 4.

As a last piece of evidence in this market regime analysis, the first two columns of Table III report forecasting performance statistics separately for recessionary and expansionary periods as determined by the National Bureau of Economic Research (NBER). This type of analysis has been relatively common in the ERP forecasting literature; see Rapach, Strauss and Zhou (2010), Henkel, Martin and Nardari (2011), Neely et al. (2014), Pettenuzzo et al. (2014), and Huang et al. (2015, 2016).

[Table III about here]

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. On the contrary, during recessionary periods (122 months in our sample), the unconstrained OLS model provides more accurate ERP forecasts, which can have important asset allocations implications for a risk-averse investor. Before looking in detail in these implications, we provide one last piece of conditional analysis on the performance of the various forecasting models, by evaluating their accuracy across different volatility regimes.

3.4. Forecasting accuracy and the volatility regime

Equity returns and volatility innovations are typically negatively correlated (among others see Ang and Bekaert 2002) with spikes in volatility typically occurring 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 be detrimental for the performance of heavily constrained models, exactly because such periods are more likely to experience negative ERP realisations. Instead, during periods of milder volatility, the constrained models are expected to outperform the unconstrained model.

The last three columns of Table III compare 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 in 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. We do acknowledge that these breakpoints are largely dependent on the sample period that is considered for our analysis and are bound to change if one uses a different sample period. By all means, 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; indicatively, the $R^2_{OOS,OLS}$ of the PTV model in the low-volatility regime is 9.98%, statistically strong at 1% level. Conversely, during periods of high volatility, the constrained generate substantially larger forecast errors compared to the unconstrained OLS model.

Overall, the findings of all parts of the current Section are in line with our hypothesis and highlight the advantages and more importantly the pitfalls of using heavily constrained models for ERP forecasting. Over the business cycle the constrained models outperform, mainly because the average ERP realisation has been positive. However, such models appear to struggle on a conditional basis. Down markets, recessions and/or high volatility periods constitute an important challenge for constrained models. This is of particular interest, because these periods are those that matter the most for a risk-averse investor, who primarily wants to hedge against unexpected increases in marginal utility. The following Section focuses in detail on the implications of investment decisions that are based on different ERP forecasting models.

4. Asset Allocation Implications

Our findings insofar can have important implications for a risk-averse investor that tactically allocates between a risky asset, which is assumed to be the equity market, and a risk-free asset. 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 her wealth that is 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. Our focus is on the evaluation of different ERP forecasting models and for that reason 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, Huang et al. 2015, Huang et al. 2016).

Typically, the large majority of academic 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:

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

Our baseline case assumes 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, we do not allow for leverage; Huang et al. (2016) 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. We evaluate these dynamics over the course of this section.

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

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

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 in order 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 in order to represent annual estimates.

Given that the focus of the paper is on the comparison between unconstrained and constrained forecasting models, 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 IV presents the economic benefit from using constrained CT and PTV forecasts based on d/p for our baseline set of parameters ($\gamma = 3$, $w_{min} = -0.5$, $w_{max} = 1$); the table reports the CER gain, the volatility of the portfolio and the Sharpe ratio of the portfolio. Panel A refers to the overall out-of-sample period, Panels B and C present sub-sample analysis across up and down markets and Panels D and E present sub-sample analysis across NBER expansions and recessions.

[Table IV about here]

The evidence is 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 in Table IV), which constitutes evidence that these models generate larger economic benefits for a risk-averse investor across the business cycle [11]. However, when studying the performance of the models conditional on the market regime (Panels B-C and D-E in Table IV), the CER gains of the constrained model appear to largely accumulate during up markets and expansionary periods. On the contrary, 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).

In the remaining part of this section, we evaluate the marginal effects of the three main parameters of the asset allocation experiment, namely, 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) assume $w_{min} = 0$ and $w_{max} = 1.5$; instead, Huang et al. (2016) assume $w_{min} = -0.5$ and $w_{max} = 1$, 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 $\gamma = 1, 3, 5$.

[11] It is important to highlight that our focus in this paper is the comparison of constrained models against the unconstrained OLS model and not the historical average, which is the typical forecasting benchmark in the literature. For the sake of completeness, the PTV model generates positive CER gains against a historical average model for our sample period. These results are available upon request from the authors.

4.1. The effect of risk aversion

Following from Equation (14), the more risk averse the investor is, 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 the risk aversion parameter: 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 $w_{max} = 1$. Figure 6 presents for all ERP forecasting models (OLS, CT and PTV) the average equity weight and the CER for the different levels of risk aversion.

[Figure 6 about here]

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's justified by the convergence of the graphs in Figure 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 stricter the minimum weight constraint is, the more likely it is that the constraint becomes binding more often. Panel A of Figure 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. The histograms of the ERP forecasts presented earlier in Figure 2 can assist in the discussion of these statistics. Furthermore, Panel B of Figure 7 presents the average equity weight and the CER for all forecasting models and different levels of permissible minimum weight.

[Figure 7 about here]

As already explained, Figure 7 shows that 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 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 becomes rarely binding for the PTV model. This justifies the straight lines in Panel B of Figure 7 for the PTV model.

Overall, the main finding from this analysis is that for an investor, who faces short-sale constraints or even worse 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 as it's justified by the convergence of the graphs in Panel B of Figure 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 because of investment mandate or just because of a broad dislike to leverage and the risks that are associated with it (Gârleanu and Pedersen 2011, Asness, Frazzini and Pedersen 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 $w_{min} = -0.5$.

As it was the case in the effect of the minimum weight, the stricter the maximum weight constraint is, the more likely it is that the constraint becomes binding more often. Figure 8 follows the structure of Figure 7 and presents in Panel A, the average amount of time (across 804 months of the out-of-sample period), that the maximum weight constraint becomes binding for the different levels of maximum weight for all ERP forecasting models (OLS, CT and PTV). Panel B of the Figure presents the average equity weight and the CER for all forecasting models and different levels of permissible maximum weight.

[Figure 8 about here]

Following from Panel A of Figure 8 and contrary to the marginal effect of the minimum weight constraint, the maximum weight constraint becomes more often binding for the PTV model, which generates on average larger ERP forecasts as documented earlier in the ERP histograms of Figure 2. Interestingly, 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 average equity allocation increases significantly, as the maximum weight constraint becomes looser (i.e. as w_{max} increases). Panel B of Figure 8 shows that the PTV model leads on average to leveraged equity allocations (i.e. average weight exceeds 1.0), when the maximum weight constraint is 1.3 or above. This leveraged investment reads to a more aggressive allocation that naturally comes along with higher volatility. This explains the relative reduction of the CER spread against the OLS unconstrained model (this spread is the *CER gain vs. OLS* defined in Equation 17), when leverage is allowed (i.e. when $w_{max} \geq 1.0$).

5. Conclusions

Constrained ERP forecasting models have been recently introduced in the academic literature, in order to alleviate the lacklustre out-of-sample forecasting performance of macroeconomic variables (Goyal and Welch 2008). Empirical analysis by Campbell and Thompson (2008) and Pettenuzzo et al. (2014) documents that such constrained models significantly improve the forecasting ability of a simple OLS regression framework over a long sample period and across the business cycle. Our focus is to investigate the benefits of these models to an asset allocator who dynamically allocates capital in the equity market; our analysis highlights the importance of conditional analysis, as opposed to an unconditional one.

Using a series of empirical tests both across time and across market states, we show that, even if the constrained models do indeed improve the forecasting performance of simple unconstrained models on an unconditional basis (across a long out-of-sample period from 1947 to 2013), the result changes aggressively on a conditional basis. During periods of high volatility and market drawdowns, the constrained forecasting models lead to large relative economic losses, compared to unconstrained models. This finding is of critical importance for tactical asset allocation decisions and raises concerns for the practical applicability of such constrained models. An asset allocator would be averse to large relative losses during turbulent periods, as this can put at risk the retention of assets under management and client base.

Seen from a different angle, our results also show that any benefit the constrained models can bring –pronounced during up markets– is substantially diminished if investors face strict investment constraints, either by mandate or regulation. Similar benefit diminution occurs when the investor is very risk averse, as the allocation to a risky asset is largely reduced.

Overall, our paper illustrates that recent academic methodological advances that have been introduced to alleviate the empirical failure of macroeconomic variables to forecast next month's ERP out-of-sample, could face a significant challenge if they were to be implemented in practice for dynamic asset allocation. Future attempts to design ERP forecasting models should carefully address the market regime dependence that we have uncovered. The state-dependent model of Huang et al. (2016) is, to our knowledge, the first attempt 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. We leave this for future research.

References

Ang, A. and Bekaert, G.: 2002, International asset allocation with regime shifts, *Review of Financial Studies* **15**(4), 1137–1187.

Ang, A. and Bekaert, G.: 2007, Stock return predictability: Is it there?, *Review of Financial Studies* **20**(3), 651–707.

Asness, C. S., Frazzini, A. and Pedersen, L. H.: 2012, Leverage aversion and risk parity, *Financial Analysts Journal* **68**(1), 47–59.

Baker, M. and Wurgler, J.: 2000, The equity share in new issues and aggregate stock returns, *Journal of Finance* **55**(5), 2219–2257.

Bollerslev, T., Tauchen, G. and Zhou, H.: 2009, Expected stock returns and variance risk premia, *Review of Financial Studies* **22**(11), 4463–4492.

Boudoukh, J., Richardson, M. and Whitelaw, R. F.: 2008, The myth of long-horizon predictability, *Review of Financial Studies* **21**(4), 1577–1605.

Breen, W., Glosten, L. R. and Jagannathan, R.: 1989, Economic significance of predictable variations in stock index returns, *Journal of Finance* **44**(5), 1177–1189.

Campbell, J. Y.: 1987, Stock returns and the term structure, *Journal of Financial Economics* **18**(2), 373–399.

Campbell, J. Y. and Cochrane, J. H.: 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* **107**(2), 205–251.

Campbell, J. Y. and Shiller, R. J.: 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* **1**(3), 195–228.

Campbell, J. Y. and Thompson, S. B.: 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* **21**(4), 1509–1531.

Campbell, J. Y. and Vuolteenaho, T.: 2004, Inflation illusion and stock prices, *American Economic Review* **94**(2), 19–23.

Campbell, J. Y. and Yogo, M.: 2006, Efficient tests of stock return predictability, *Journal of Financial Economics* **81**(1), 27–60.

Clark, T. E. and West, K. D.: 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* **138**(1), 291–311.

Cochrane, J. H.: 2011, Presidential address: Discount rates, *Journal of Finance* **66**(4), 1047–1108.

Cooper, I. and Priestley, R.: 2009, Time-varying risk premiums and the output gap, *Review of Financial Studies* **22**(7), 2801–2833.

Fama, E. F.: 1986, Term premiums and default premiums in money markets, *Journal of Financial Economics* **17**(1), 175–196.

Fama, E. F. and French, K. R.: 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* **22**(1), 3–25.

Fama, E. F. and Schwert, G. W.: 1977, Asset returns and inflation, *Journal of Financial Economics* **5**(2), 115–146.

Ferreira, M. A. and Santa-Clara, P.: 2011, Forecasting stock market returns: The sum of the parts is more than the whole, *Journal of Financial Economics* **100**(3), 514–537.

French, K. R., Schwert, G. W. and Stambaugh, R. F.: 1987, Expected stock returns and volatility, *Journal of Financial Economics* **19**(1), 3–29.

Gârleanu, N. and Pedersen, L. H.: 2011, Margin-based asset pricing and deviations from the law of one price, *Review of Financial Studies* **24**(6), 1980–2022.

Goyal, A. and Welch, I.: 2003, Predicting the equity premium with dividend ratios, *Management Science* **49**(5), 639–654.

Goyal, A. and Welch, I.: 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* **21**(4), 1455–1508.

Henkel, S. J., Martin, J. S. and Nardari, F.: 2011, Time-varying short-horizon predictability, *Journal of Financial Economics* **99**(3), 560–580.

Huang, D., Jiang, F., Tu, J. and Zhou, G.: 2015, Investor sentiment aligned: a powerful predictor of stock returns, *Review of Financial Studies* **28**(3), 791–837.

Huang, D., Jiang, F., Tu, J. and Zhou, G.: 2016, Forecasting stock returns during good and bad times, *SSRN eLibrary* .

Keim, D. B. and Stambaugh, R. F.: 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* **17**(2), 357–390.

Kothari, S. P. and Shanken, J.: 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* **44**(2), 169–203.

Lettau, M. and Ludvigson, S.: 2001, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* **56**(3), 815–849.

Li, J. and Yu, J.: 2012, Investor attention, psychological anchors, and stock return predictability, *Journal of Financial Economics* **104**(2), 401–419.

Li, Y., Ng, D. T. and Swaminathan, B.: 2013, Predicting market returns using aggregate implied cost of capital, *Journal of Financial Economics* **110**(2), 419–436.

Møller, S. V. and Rangvid, J.: 2015, End-of-the-year economic growth and time-varying expected returns, *Journal of Financial Economics* **115**(1), 136–154.

Neely, C. J., Rapach, D. E., Tu, J. and Zhou, G.: 2014, Forecasting the equity risk premium: the role of technical indicators, *Management Science* **60**(7), 1772–1791.

Pettenuzzo, D., Timmermann, A. and Valkanov, R.: 2014, Forecasting stock returns under economic constraints, *Journal of Financial Economics* **114**(3), 517–553.

Pontiff, J. and Schall, L. D.: 1998, Book-to-market ratios as predictors of market returns, *Journal of Financial Economics* **49**(2), 141–160.

Rapach, D. E., Ringgenberg, M. C. and Zhou, G.: 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* **121**(1), 46–65.

Rapach, D. E., Strauss, J. K. and Zhou, G.: 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* **23**(2), 821–862.

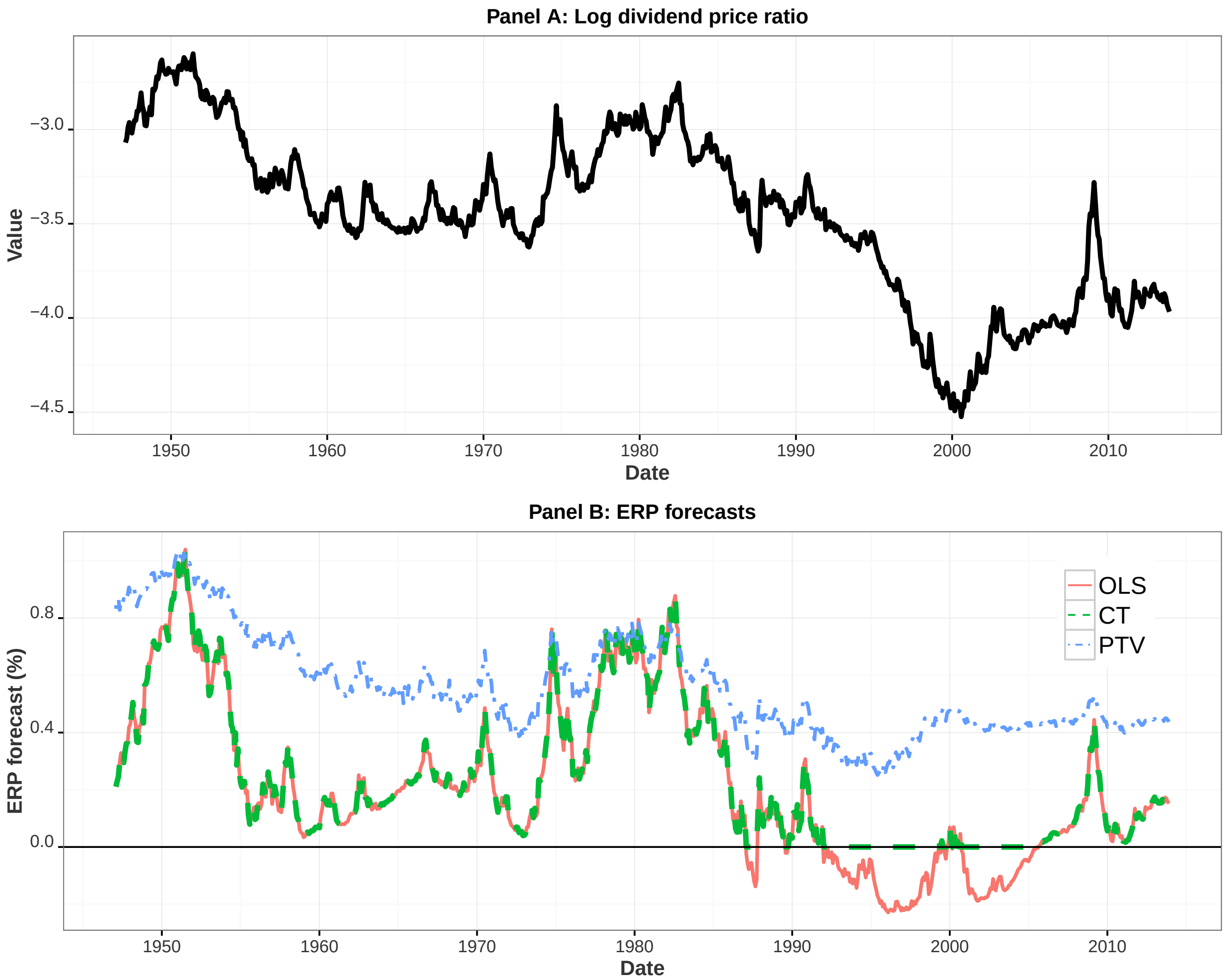


Figure 1: *ERP forecasts using the dividend-price ratio*
The figure presents the historical values of the dividend-price ratio (difference between the log of the 12-month moving sums of dividends and the log of price of the S&P 500 index) in Panel A and the ERP forecasts in Panel B. Three different forecasting models are used: the baseline OLS regression model, the Campbell and Thompson (2008, CT) truncation model and the Pettenuzzo, Timmermann and Valkanov (2014, 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.

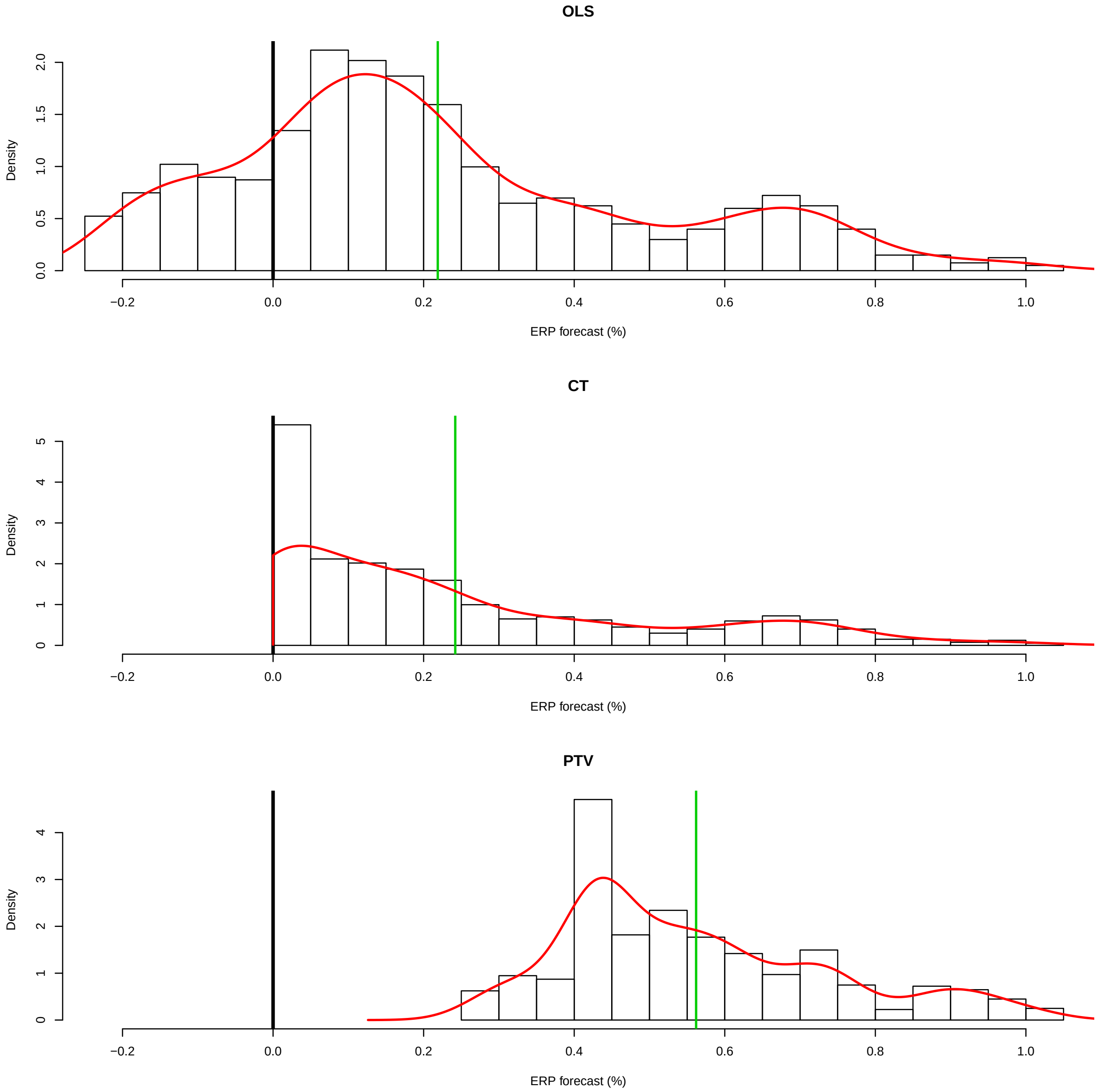


Figure 2: *Histograms of ERP Forecasts using d/p as predictor*
The figure presents the histograms of ERP forecasts using the dividend-price ratio (d/p) as predictor. Three different forecasting models are used: the baseline OLS regression model, the Campbell and Thompson (2008, CT) truncation model and the Pettenuzzo, Timmermann and Valkanov (2014, 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 green vertical line.

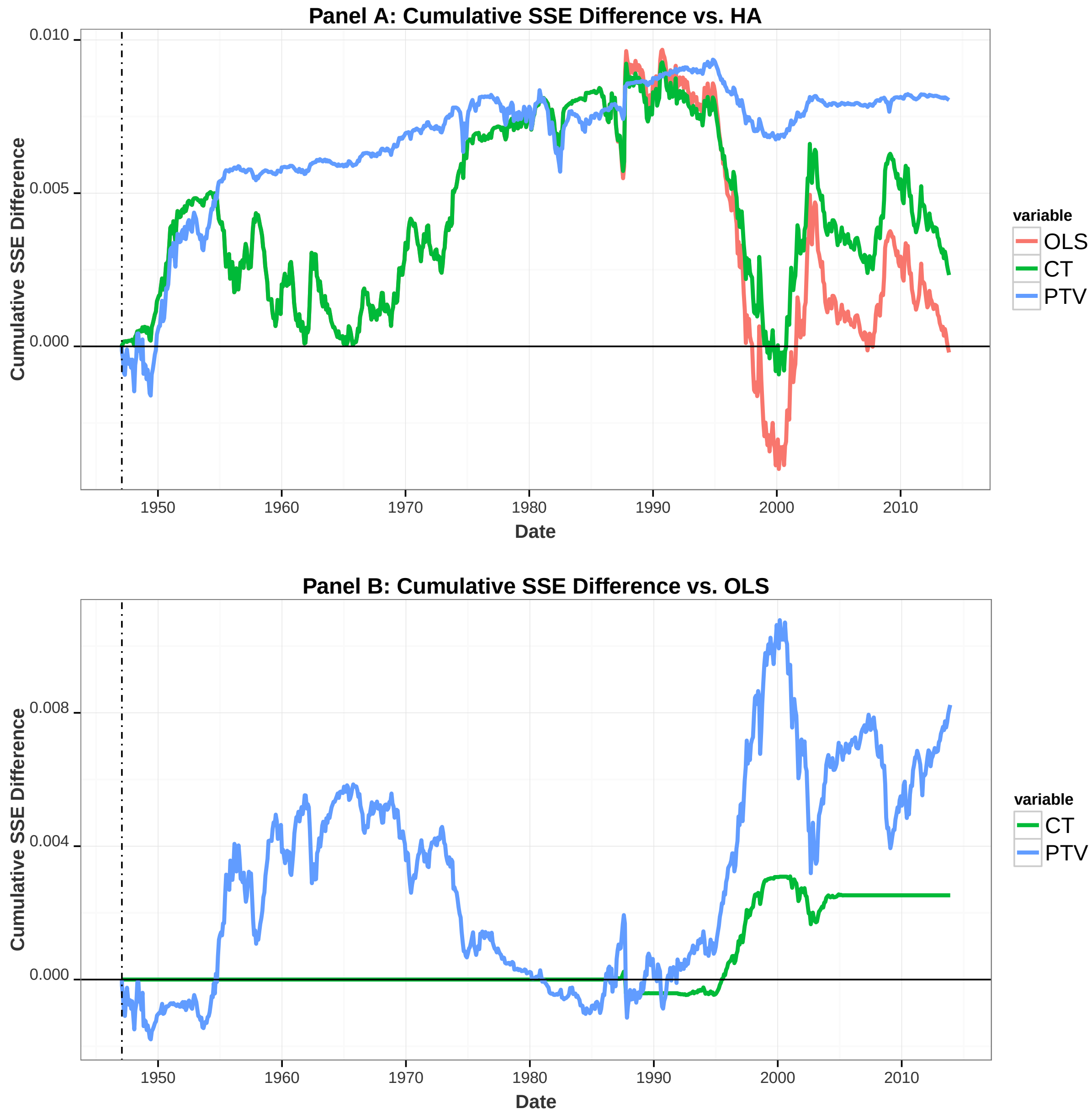


Figure 3: *Cumulative sum of squared forecast errors*
The figure presents the difference in the cumulative sum of squared forecast errors (*DCSSE*) between the three different forecasting models (the baseline OLS regression model, the Campbell and Thompson (2008, CT) truncation model and the Pettenuzzo, Timmermann and Valkanov (2014, 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 (d/p). The initial training period is from January 1927 to December 1946 and the out-of-sample period is from January 1947 to December 2013.

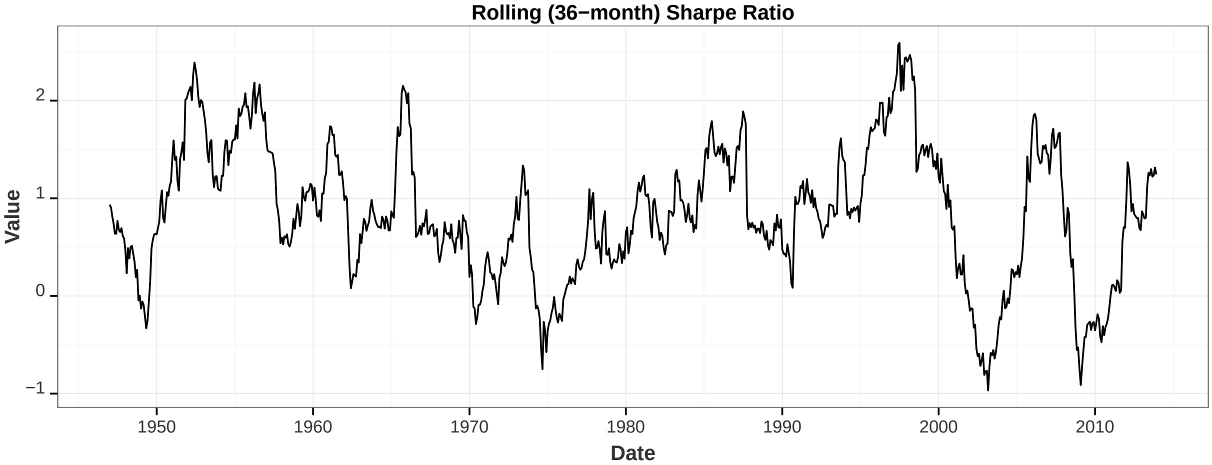


Figure 4: *Rolling (36-month) Sharpe ratio of the S&P500 index*
The figure presents a rolling Sharpe ratio calculation at the end of each month for the S&P500 index using a window of 36 months. The sample period is from January 1947 to December 2013.

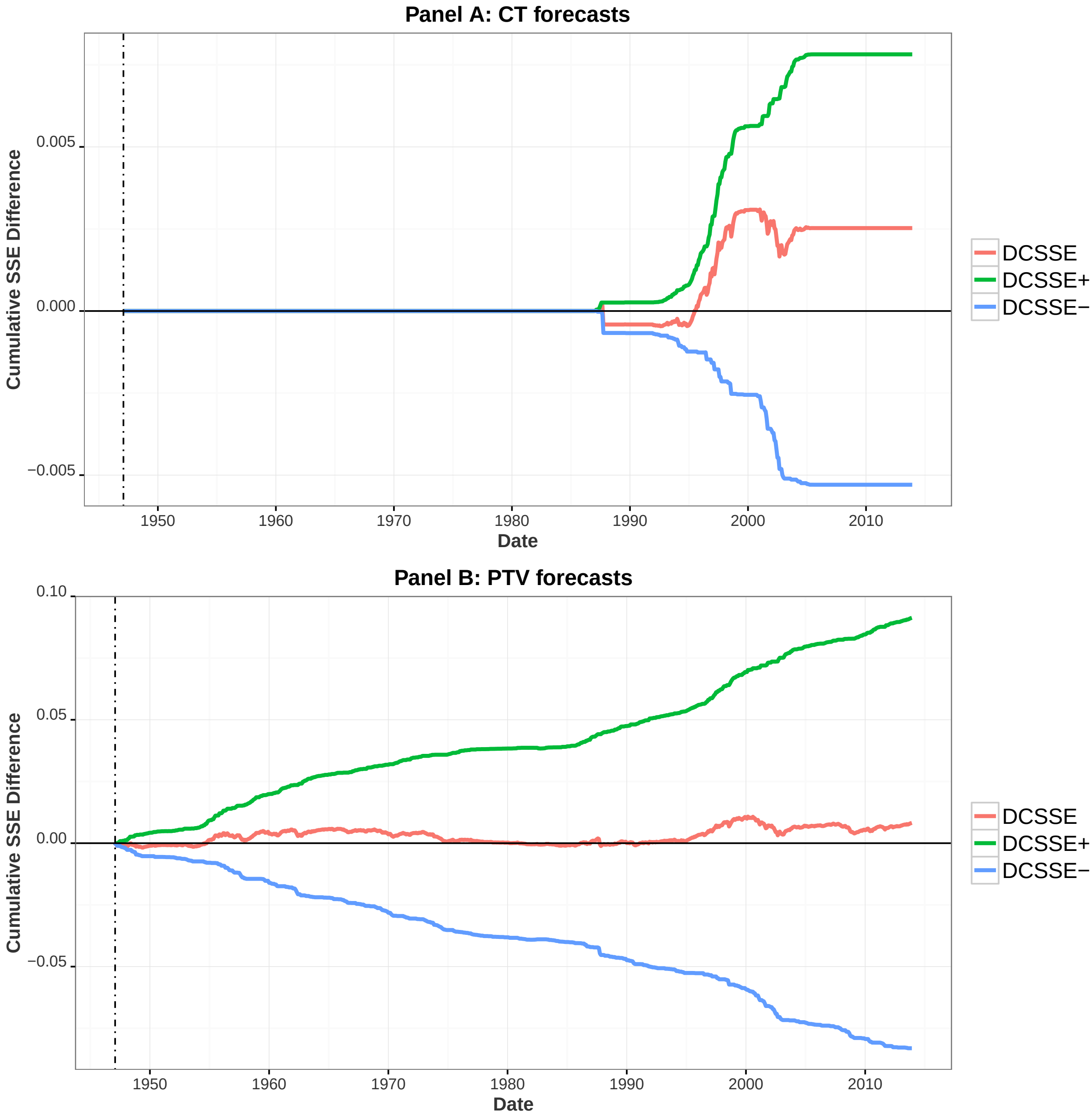


Figure 5: *Constrained versus Unconstrained models in up and down markets*
The figure presents the difference in the cumulative sum of squared forecast errors (*DCSSE*) between the unconstrained OLS model and the constrained models; the Campbell and Thompson (2008, CT) truncation model in Panel A and the Pettenuzzo, Timmermann and Valkanov (2014, 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.

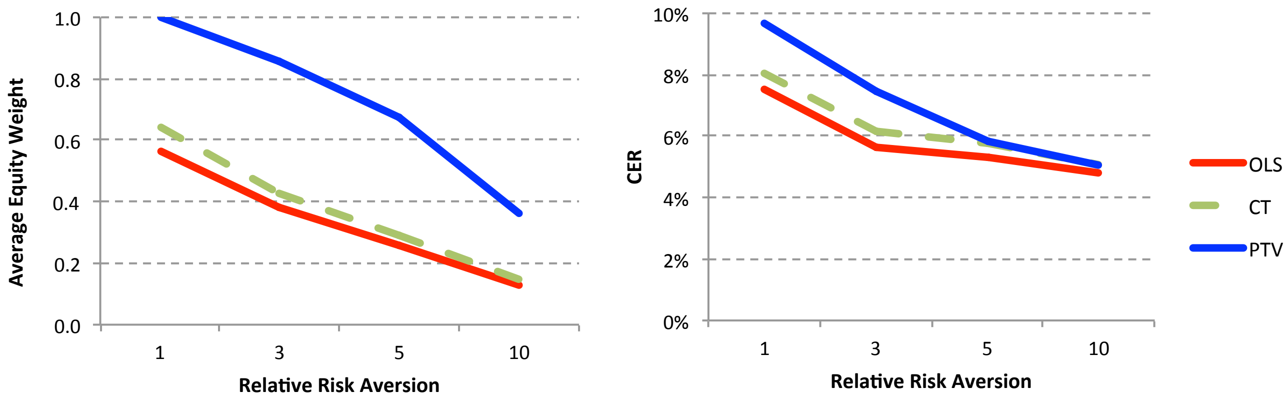
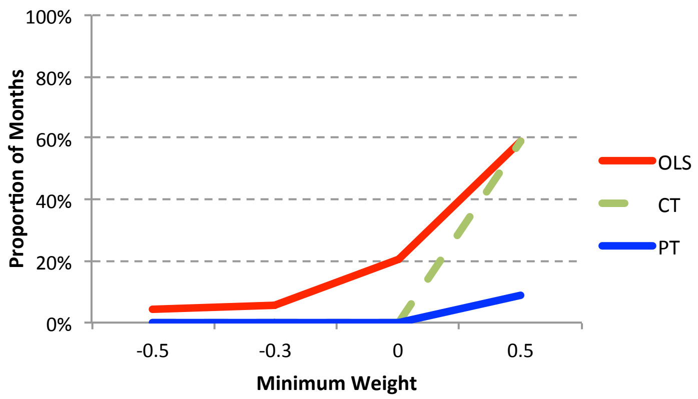


Figure 6: *The Effect of Risk Aversion*

The figure presents the marginal impact on the average equity weight (left pane) and on the CER (right pane) 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 Campbell and Thompson (2008, CT) model and the constrained Pettenuzzo, Timmermann and Valkanov (2014, 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.

Panel A: How often does the minimum weight constraint become binding?



Panel B: Average weight and CER for different levels of minimum weight ($\gamma = 3$ and $w_{max} = 1$)

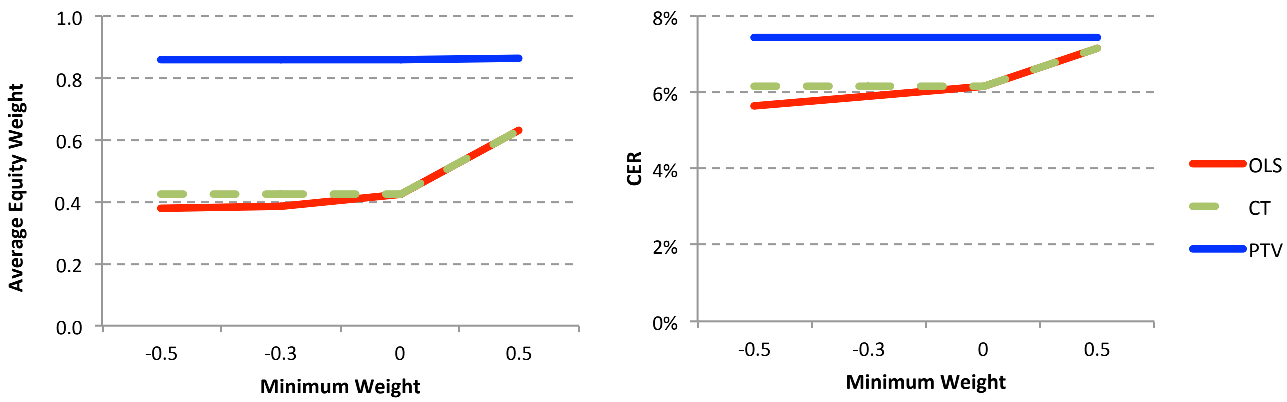
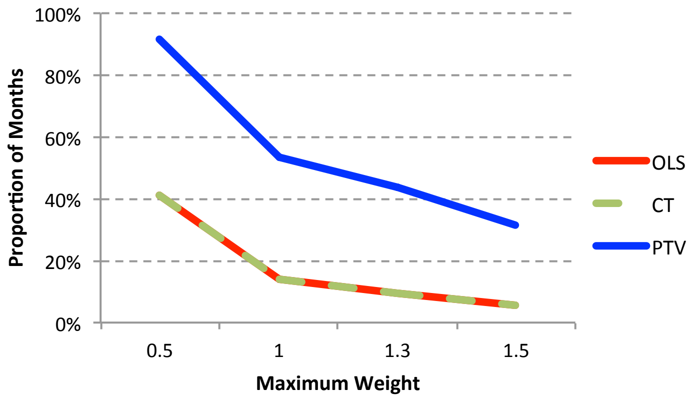


Figure 7: *The Effect of Minimum Weight Constraint*
The figure presents in Panel A the proportion of months that different levels of the minimum weight constraint become binding, when using out-of-sample ERP forecasts that are generated by the dividend-price ratio using the unconstrained OLS model, the constrained Campbell and Thompson (2008, CT) model and the constrained Pettenuzzo, Timmermann and Valkanov (2014, PTV) model. Panel B presents the marginal impact on the average equity weight (left pane) and on the CER (right pane) 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.

Panel A: How often does the maximum weight constraint become binding?



Panel B: Average weight and CER for different levels of maximum weight ($\gamma = 3$ and $w_{min} = -0.5$)

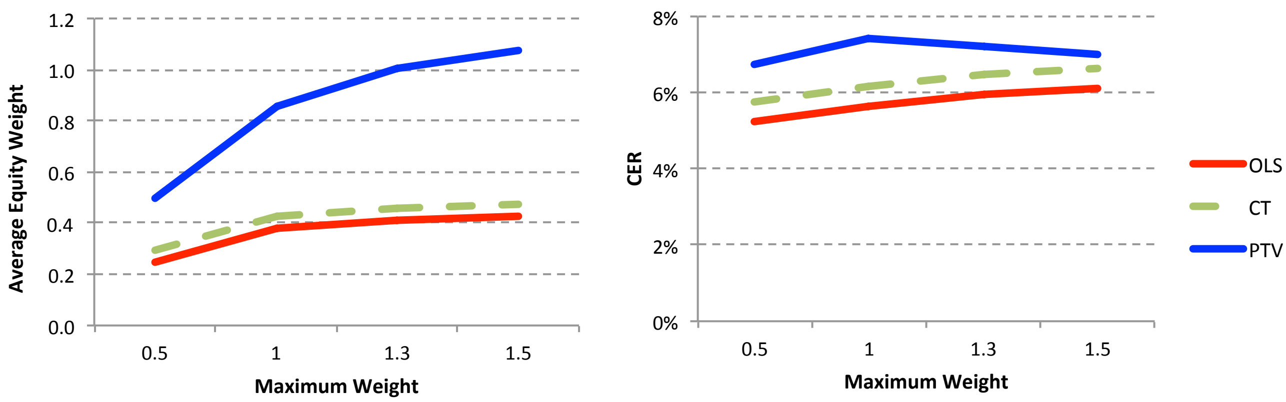


Figure 8: *The Effect of Maximum Weight Constraint*
The figure presents in Panel A the proportion of months that different levels of the maximum weight constraint become binding, when using out-of-sample ERP forecasts that are generated by the dividend-price ratio using the unconstrained OLS model, the constrained Campbell and Thompson (2008, CT) model and the constrained Pettenuzzo, Timmermann and Valkanov (2014, PTV) model. Panel B presents the marginal impact on the average equity weight (left pane) and on the CER (right pane) 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.

		All	Down	Up
	# months	804	328	476
	OLS	4.227	4.746	3.828
RMSFE x 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%

Table I: *Out-of-sample performance*

The table compares the out-of-sample forecasting ability of the constrained Campbell and Thompson (2008, CT) and Pettenuzzo, Timmermann and Valkanov (2014, PTV) models against the unconstrained OLS model when the dividend-price ratio (d/p) is used as the predictor variable across the full sample ("All") and across months with a positive ERP realisation ("up") and months with a negative ERP realisation ("down"). The out-of-sample period is January 1947 to December 2013. The constrained models are compared against the unconstrained OLS model by the means of (a) RMSFE: the root mean square forecast error, (b) $R^2_{OOS,OLS}$: the out-of-sample (OOS) R^2 with respect to the unconstrained OLS model, which is defined as the proportional decrease in the mean squared error of the unconstrained model when a constrained model is employed and (c) the proportion of months that a constrained model generates a smaller absolute forecast than the unconstrained OLS. The null hypothesis $H_0 : R^2_{OOS,OLS} \leq 0$ is evaluated against the one-sided alternative hypothesis hypothesis $H_A : R^2_{OOS,OLS} > 0$ based on the Clark and West (2007) MSFE-*adjusted* statistic; statistical significance is denoted by *, **, *** for significance at 10%, 5% and 1% levels respectively.

		NL	NS	PS	PL
	# months	164	164	238	238
ERP realisations	min	-25.0%	-2.4%	0.0%	2.9%
	max	-2.4%	0.0%	2.9%	15.0%
RMSFE x 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%

Table II: *Forecasting across market regimes*
The table presents an event study comparing the out-of-sample forecasting ability of the constrained Campbell and Thompson (2008, CT) and Pettenuzzo, Timmermann and Valkanov (2014, PTV) models against the unconstrained OLS model when the dividend-price ratio (d/p) 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 realisation (476 months) and months with a negative EPR realisation (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 between the various market regimes. The out-of-sample period is from January 1947 to December 2013. The constrained models are compared against the unconstrained OLS model separately across each regime by the means of (a) RMSFE: the root mean square forecast error, (b) $R^2_{OOS,OLS}$: the out-of-sample (OOS) R^2 with respect to the unconstrained OLS model, which is defined as the proportional decrease in the mean squared error of the unconstrained model when a constrained model is employed and (c) the proportion of months that a constrained model generates a smaller absolute forecast than the unconstrained OLS. The null hypothesis $H_0 : R^2_{OOS,OLS} \leq 0$ is evaluated against the one-sided alternative hypothesis hypothesis $H_A : R^2_{OOS,OLS} > 0$ based on the Clark and West (2007) MSFE-*adjusted* statistic; statistical significance is denoted by *, **, *** for significance at 10%, 5% and 1% levels respectively.

		Business Cycles		Volatility Regimes		
		Rec	Exp	Low	Med	High
	# months	122	682	268	268	268
	OLS	5.774	3.885	2.848	3.609	5.696
RMSFE x 100	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	CT	2.46%	15.12%	11.19%	13.11%	15.30%
smaller error vs. OLS	PTV	43.44%	57.71%	66.04%	54.68%	45.90%

Table III: *ERP Forecasting across business cycles and volatility regimes*
The table presents an event study comparing the out-of-sample forecasting ability of the constrained Campbell and Thompson (2008, CT) and Pettenuzzo, Timmermann and Valkanov (2014, PTV) models against the unconstrained OLS model when the dividend-price ratio (d/p) 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 realised volatility of S&P 500 index (sum of daily logarithmic returns). The out-of-sample period is January 1947 to December 2013. The constrained models are compared against the unconstrained OLS model separately across each regime by the means of (a) RMSFE: the root mean square forecast error, (b) $R^2_{OOS,OLS}$: the out-of-sample (OOS) R^2 with respect to the unconstrained OLS model, which is defined as the proportional decrease in the mean squared error of the unconstrained model when a constrained model is employed and (c) the proportion of months that a constrained model generates a smaller absolute forecast than the unconstrained OLS. The null hypothesis $H_0 : R^2_{OOS,OLS} \leq 0$ is evaluated against the one-sided alternative hypothesis hypothesis $H_A : R^2_{OOS,OLS} > 0$ based on the Clark and West (2007) MSFE-*adjusted* statistic; statistical significance is denoted by *, **, *** for significance at 10%, 5% and 1% levels respectively.

Baseline parameter values: $\gamma = 3$, $w_{min} = -0.5$, $w_{max} = 1$			
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

Table IV: *Asset Allocation Results - Baseline Scenario*
The table reports the economic benefit for a mean-variance investor with a risk aversion of three ($\gamma = 3$) that uses constrained Campbell and Thompson (2008, CT) and Pettenuzzo, Timmermann and Valkanov (2014, PTV) models in order to generate ERP forecasts using the dividend-price ratio (d/p) as the predictor variable. The reported statistics are: the Certainty Equivalent Returns (CER) gain against using an unconstrained OLS model, the annualised 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.