A Predictive System with Heteroscedastic Expected Returns and Economic Constraints*

Maxime Bonelli† and Daniel Mantilla-García‡

March 26, 2015

Abstract

We propose a variation of a predictive system that incorporates two additional economically motivated assumptions about the dynamics of expected market returns, namely 1) a time-varying conditional volatility, and 2) their non-negativity. The modified system without predictors can explain the well documented countercyclicality of the dividend-price ratio's predictive power, and can produce significantly lower out-of-sample forecast errors than the historical mean, as well as some improvement compared to the original system. Furthermore, the Bayesian estimation of the model indicates that the persistence parameter of expected returns in the modified system has been declining over the last couple of decades in tandem with the sample autocorrelation of realized returns.

Keywords: return predictability, economic constraints, dividend-price ratio, Kalman filter, Bayesian analysis.

JEL classification: C58, G00, G17, C11.

---

*A previous version of this paper circulated with the title "Predictive Systems under Economic Constraints". We would like to thank Mireille Bossy for the proof of Proposition 1 in Appendix B and for valuable comments, Abraham Lioui and Georges Hübner for very insightful discussions and ideas, and Inria Sophia Antipolis - Méditerranée research centre for allowing us to use one of their calculating servers. Financial support from Koris International is gratefully acknowledged.

†Contact author, PhD student at Inria Sophia Antipolis - Méditerranée research centre, TOSCA team and Research engineer at Koris International. **Email**: maxime.bonelli@inria.fr, **Address**: Koris International, Parc Technologique de Sophia-Antipolis, 200 avenue de Roumanille, Immeuble Néri, 06410 Biot - France, and **Phone**: + 33 (0)4 88 72 88 32.

‡Research associate at Edhec-Risk Institute and Head of Research at Optimal Asset Management, 171 Main St # 298 Los Altos, CA 94022 USA. **Email**: daniel.mantilla@edhec.edu.

1 Introduction

The estimation of expected stock market returns[1], is a central issue in financial economics from both theoretical and applied standpoints. The classic random walk paradigm assumes expected market returns are constant implying nil return predictability, while the increasingly accepted time-varying expected return paradigm implies *some* predictability must exist.

The question of return predictability of the stock market has mostly been addressed within the linear predictive regression framework (see for instance Keim and Stambaugh, 1986; Stambaugh, 1986; Ferson and Harvey, 1991; Pesaran and Timmermann, 1995; Stambaugh, 1999; Goyal and Welch, 2008; Lettau and van Nieuwerburgh, 2008; Rapach, Strauss, and Zhou, 2010; Dangl and Halling, 2012). Several economic predictors have been investigated, including the dividend yield (Fama and French, 1988; Campbell and Shiller, 1988; Goyal and Welch, 2003; Ang and Bekaert, 2007; Cochrane, 2008), the interest rates (Campbell, 1987), the term and default spreads (Campbell, 1987; Fama and French, 1989), and the consumption-wealth ratio (Lettau and Ludvigson, 2001).

The regression approach has some limits, among which the fact that it assumes a perfect linear relationship between the predictor(s) and expected returns. Recently Pástor and Stambaugh (2009) introduced a *predictive system* allowing for imperfect predictors, which is a richer environment than the standard predictive regression to analyze the interactions amongst realized returns, expected returns and predictors.

Pástor and Stambaugh (2009)'s predictive system assumes that the unobservable expected return process follows an Autoregressive process (AR) of order 1, which has a constant conditional variance. Motivated by empirical observations and economic theory, we propose to use instead a discretization of a Cox, Ingersoll Jr, and Ross (1985) (CIR) process to model the dynamics of expected market returns, which is a very parsimonious departure from the original system, with no need of additional parameters. Unlike the initial AR system, the CIR dynamics induce a continuously changing conditional variance for expected returns that increases during market downturns (assuming expected market returns are countercyclical). The motivation for a time-varying variance in expected returns follows from the combination of two pervasive empirical facts in stock markets: 1) during economic and market downturns the variance of realized returns increases (see

[1] Unless explicitly stated otherwise, we use the term *expected returns* as a shorthand for *expected excess returns of the stock market* over the risk-free rate or *equity risk premium.* Similarly, *realized returns* stands for *realized excess market returns.*

Schwert, 1989; Hamilton and Lin, 1996; Ait-Sahalia and Kimmel, 2007), and 2) the predictability of returns also increases during economic recessions (see Rapach et al., 2010; Henkel, Martin, and Nardari, 2011). Return predictability is measured as the fraction of the variance in realized returns explained by variations in expected returns, i.e. the R^2 of the regression of realized returns on expected returns. It follows that the variance of expected returns must increase during economic recessions to compensate for the variance rise of realized returns. Furthermore, uncertainty about economic prospects also increases during recessions, which can (arguably) be translated as an increase in the variance of expected returns.

This paper does not investigate the predictive power of a given predictor, but focuses instead on the implications arising from a modified interaction between past returns and expected returns, due to the new dynamics of the latter. Hence our study concentrates on the system without predictors, in order to present results that remain valid regardless of the future choice of eventual predictors[2]. However, as Figure 3 illustrates, we find that the conditional heteroscedasticity of expected returns in the CIR system without predictors reproduces the countercyclical predictability of the dividend yield in the predictive regression documented in former studies such as Rapach et al. (2010), Mantilla-Garcia and Vaidyanathan (2011), and Henkel et al. (2011).

Another difference with Pástor and Stambaugh (2009)'s original system is that the modified system also implies variations in the conditional variance of realized returns. If expected returns are countercyclical, then the changes in the conditional variance of realized returns produced by the modified system have a positive correlation with expected returns and a negative one with realized returns, which is consistent with the empirical observations in Ait-Sahalia and Kimmel (2007).

Moreover, the CIR-based expected return process has a negligible probability of being negative. This additional feature is compatible with economic intuition: risk-averse investors would not hold stocks if the equity premium was negative[3]. In that sense, Campbell and Thompson (2008) and Pettenuzzo, Timmermann, and Valkanov (2014) have shown that the economically motivated belief of positive equity expected excess returns can be used to improve the robustness of return forecasts within the predictive regression framework. In our version of the system, the positivity condition is simply a

[2]In Appendix C we present the system in its general version with external predictors.

[3]Merton (1980) estimates instantaneous expected return on the market and concludes that: *"in estimating models of the expected market return, the non-negativity restriction of the expected excess return should be explicitly included as part of the specification"* (Merton, 1980, p. 323).

result of the proposed dynamics of the expected returns process. This structural approach allows us to explore the theoretical implications of this prior belief.

In our empirical analysis we find that the predictive system can produce return forecasts out-of-sample that are significantly better than the historical average, as well as some forecast improvements relative to the original system, if economically motivated priors are used for the parameters of the system. The results indicate that the modified system produces better forecasts under the priors that expected returns are countercyclical and display a relatively low variance. On the other hand, the estimation of the persistence parameter of expected returns in the CIR system is less close to 1 than in the AR system. Furthermore its value in the CIR system has been declining over the last couple of decades in tandem with the sample autocorrelation of realized returns, which contrasts with the higher and stable estimated value in the original AR system.

2 A predictive system with heteroscedastic and positive expected returns

In this section we present a short summary of Pástor and Stambaugh (2009) predictive system and then introduce the modified version of the system along with its theoretical implications.

2.1 Pástor and Stambaugh (2009)'s predictive system

Pástor and Stambaugh (2009) assume the following discrete dynamics for the realized return r at time $t+1$:

$$r_{t+1} = \mu_t + u_{t+1}, \tag{1}$$

where the innovation u_{t+1} is the "unexpected return". The *unobservable* expected return μ, follows a first-order autoregressive AR(1) process:

$$\mu_{t+1} = (1-\beta)E_r + \beta\mu_t + w_{t+1}, \tag{2}$$

where E_r denotes the unconditional expectation of r, which is equal to the unconditional expectation of μ and is constant in time; β is a constant persistence parameter assumed to be within $(0, 1)$ so that μ is stationary, and w_{t+1} is the innovation in the expected return. We will refer later to this predictive system introduced by Pástor and Stambaugh (2009) as the AR (autoregressive) system.

Besides, they also consider a set of stationary (observable) predictors x_t following a first-order vector autoregressive VAR(1) process (a standard assumption in the predictability literature),

$$x_{t+1} = (I - A)E_x + Ax_t + v_{t+1}, \tag{3}$$

where E_x is the unconditional expectation of x, A is a matrix with suitable dimensions containing the autoregressive coefficients and with eigenvalues lying inside the unit circle, and v is gaussian noise. Furthermore, the three innovation processes above are assumed to be correlated white-noise, independent and identically distributed across t as,

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{uw} \\ \sigma_{vu} & \Sigma_{vv} & \sigma_{vw} \\ \sigma_{wu} & \sigma_{wv} & \sigma_w^2 \end{bmatrix} \right). \tag{4}$$

Denote the covariance matrix in (4) as Σ. Notice that the interaction between predictors and expected returns happens trough the correlation between their corresponding innovations v and w. While in the standard predictive regression the correlation between the predictor and expected returns is assumed to be perfect, as $\mu_t = a + b'x_t$, for constant a and b with suitable dimensions, the predictive system (1), (2), (3), (4) allows for "imperfect predictors" presenting a correlation with expected returns lower than 1 in magnitude. This "imperfect correlation" implies that the estimated expected return depends on past returns. Pástor and Stambaugh (2009) showed that the standard predictive regression is a particular case of the predictive system in which the correlation between the innovation in the predictor x and innovation in μ is assumed to be perfect, i.e., $\rho_{vw} = \pm 1$, and the autoregressive coefficient of μ and x are equal, e.g., $\beta = A$ if we consider one predictor.

Another distinctive characteristic of the predictive system is the presence of a correlation between the innovation in μ and the innovation in r, i.e., ρ_{uw}. This correlation has an impact on the relation between expected returns and past realized returns, and on the relative importance of what Pástor and Stambaugh (2009) called the *level effect* and the *change effect* in the system. The level effect captures the procyclicality of expected returns, i.e., the extent to which observing relatively higher (lower) realized returns is a signal of higher (lower) expected returns, while the change effect captures the extent to which observing relatively higher (lower) realized returns is a signal of lower (higher) expected returns (countercyclicality). For the change effect to dominate, ρ_{uw} must be sufficiently negative. If the change effect prevails then expected returns are countercyclical. Pástor and Stambaugh (2009) argue that the change effect should dominate the level effect and other studies such as Campbell (1991), Campbell and Ammer (1993) and Binsbergen, Jules, and Koijen (2010) point in the same direction.

Equations (1), (2), (3) and (4) constitute a state-space model[4] in which $\mathbb{E}(\mu_t|D_t) = \mathbb{E}(r_{t+1}|D_t)$, where D_t denotes the information set available at time t of observable quantities r and x. Hence, a linear Kalman filter can be used to estimate the unobservable expected return process[5] μ.

In the Bayesian empirical analysis of the predictive system, Pástor and Stambaugh (2009) used prior distributions of the input parameters of the system reflecting "*the prior belief that the conditional expected return μ_t is stable and persistent*" (Pástor and Stambaugh, 2009, p. 1606). To capture the belief that μ_t is stable, they imposed a prior that the predictive R^2 from the regression of r_{t+1} on μ_t is not very large, which is equivalent to the belief that the total variance of μ_t is not very large relative to the variance of realized returns. To capture the belief that μ_t is persistent, they impose a prior that β, the slope of the AR(1) process for μ_t, is smaller than one but not by much.

2.2 The CIR predictive system

Empirical evidence by Henkel et al. (2011) and others shows that return predictability is markedly countercyclical, i.e., it is much stronger during economic recessions. Higher predictability means that the percentage of the variance of realized returns explained by variations in expected returns is higher. Hence, that evidence suggests that the variance of expected returns should increase more during recessions than the variance of realized returns. Furthermore, standard equilibrium models with risk-averse investors predict positive expected returns (see for instance Merton, 1980, 1993).

The AR(1) process used in the original predictive system of Pástor and Stambaugh (2009) to model expected returns does allow negative values and, more importantly, its conditional variance is constant over time. We propose a very parsimonious departure from the original system that does not introduce any additional parameters, but modi-

[4]Other studies on return predictability using state-space models include Conrad and Kaul (1988); Lamoureux and Zhou (1996); Ang and Piazzesi (2003); Brandt and Kang (2004); Duffee (2007); Rytchkov (2012).

[5]Pástor and Stambaugh (2012) presents an alternative state-space representation of the predictive system where r_t and x_t follow a VAR process with an unobserved additional predictor but as they explain this alternative representation is well suited for exploring the role of predictor imperfection which is not our aim in this paper.

fies the governing equation of the expected return process in a way that integrates the aforementioned economically motivated features.

Suppose that the true unobservable process μ follows a Cox et al. (1985) (CIR) model, which is a continuous-time mean-reverting process given by the Stochastic Differential Equation (SDE):

$$d\mu_t = \kappa(\theta - \mu_t)dt + \sigma\sqrt{\mu_t}dW_t, \tag{5}$$

where the constant κ is the speed of mean reversion, θ the long-term level of μ, σ the standard deviation of the diffusion term and W is a standard Brownian motion. The CIR process rules out negative values of μ if its parameters satisfy the condition $\kappa\theta \geq \frac{\sigma^2}{2}$, along with $\mu_0 > 0$ (see Feller, 1951). Furthermore, the diffusion factor, $\sigma\sqrt{\mu_t}$, induces a process with a level dependent time-varying volatility that increases whenever the level of μ increases. Thus, if the dynamics of μ are countercyclical, then its variance as well, everything else equal.

The CIR's SDE does not have an explicit closed-form solution[6] for μ. Thus, in order to develop the economic intuition and implications of the model, and to remain in the discrete-time framework developed by Pástor and Stambaugh (2009), we work with a direct Euler discretization of equation (5) under the assumption that the former realization of μ_t is positive, that is:

$$\mu_{t+1} = (1-\beta)E_r + \beta\mu_t + \sqrt{\mu_t}w_{t+1} \quad \text{given that } \mu_t \geq 0, \tag{6}$$

where $E_r = \theta$ is the constant long-term mean, $\beta = (1 - \kappa\Delta t)$ is the auto-regressive constant, w is gaussian innovation with variance $\sigma_w^2 = \sigma^2\Delta t$ and Δt is the time step chosen in the discretization, i.e., the elapsed time between t and $t+1$ in the time series notation.

Appendix A and B present the mild technical conditions needed for the discretization (6) to have a negligible probability to yield negative values for μ. In appendix C we derive the expressions of the extended Kalman Filter algorithm for a general function $g(\mu_t)$ on the diffusion term, instead of the particular case $g(.) = \sqrt{.}$ of equation (6).[7]

Note that Pástor and Stambaugh (2009)'s AR(1) system is nested in the general system derived in the Appendix with $g(.) = 1$. Also notice that the assumption that

[6]However the probability density of the solution of a CIR's SDE can be computed in closed form.

[7]The derivation in appendix C includes the case $g(.) = \sqrt{|.|}$, which is well defined for $\mu_t < 0$, unlike equation (6). However, for simplicity of exposure we develop the economic intuition of the model in the simpler case of equation (6) (refer to appendix A for a detailed discussion).

equity expected returns are unlikely to be negative is introduced by a structural change in the model instead of correcting the model outputs ex-post as in Campbell and Thompson (2008)[8] or imposing prior constraints on the input parameters as done by Pettenuzzo et al. (2014) in the predictive regression.

Our (full) version of the predictive system is similar to Pástor and Stambaugh (2009)'s AR(1) system discussed above but uses the state equation (6) for μ instead of equation (2), while the realized return and predictors equations (1) and (3) as well as the multivariate gaussian distribution assumption for the innovations in equation (4) are kept the same. We refer to this new model as the CIR system in the rest of the paper.

We are not aiming to analyze the predictive power of a particular predictor, thus we focus our study on the implications of the modified equation for μ, using the predictive system without predictors, i.e., equations (1) and (6), as well as joint distribution of innovations $[u_t \ w_t]'$ in (4). In that sense, our analysis remains valid regardless of the eventual predictors that could be integrated into the system, as our results concern the relation between r and μ and are not dependent of a given predictor. In Appendix C we present the equations of the general system including the possible interaction with external predictors. Furthermore, in section 2.5 we discuss the relationship of the expected returns filtered from the system without predictors (that uses only past realized returns) to estimate μ with the outputs of the standard predictive regression using the dividend yield as the predictor and find a remarkable relationship between the two estimates of μ, when ρ_{uw} is assumed to be sufficiently negative.

In what follows we explore the theoretical implications of the CIR predictive system and perform an empirical analysis using quarterly returns of the value-weighted index of all NYSE, Amex, and Nasdaq stocks in excess of the quarterly return on 1-month T-bills obtained from the Center for Research in Security Prices (CRSP). Following Pástor and Stambaugh (2009) and other studies we begin our sample in 1952-Q1 after the Fed was allowed to pursue an independent monetary policy. Our sample ends in 2012-Q4. As mentioned above, in section 2.5 we compare the outputs of the system with a predictive regression using the dividend-price ratio as the predictor variable. The latter is computed as the sum of total dividends paid over the last 12 months divided by the current price, using monthly stock returns with and without dividend on the value-weighted index from

[8]Campbell and Thompson (2008) show that the predictive regression subject to economic constraints (including non-negativity of conditional mean of the equity premium) performs better than using the unrestricted regression.

CRSP.

2.3 Implications on R^2 and the change and level effects

Assuming μ_t follows the dynamics in equation (6) has several key implications, including the way in which past realized returns impact changes in expected returns. The implications follow from the mechanics of CIR-type processes[9], in which the diffusion term of μ becomes negligible and its mean reverting strength preponderant whenever μ approaches zero. To see this, notice that from equation (6), it follows that,

$$\mu_{t+1} - \mu_t = (1-\beta)(E_r - \mu_t) + \sqrt{\mu_t} w_{t+1}. \tag{7}$$

Hence, whenever μ_t approaches zero, its diffusion term as well, making μ_{t+1} much more likely to increase driven by the mean-reversion term $(1-\beta)(E_r - \mu_t)$. Notice as well that the speed of mean reversion decreases with the level of the persistence parameter β, thus being two competing effects on the dynamics of expected returns.

Pástor and Stambaugh (2009) explore the interesting temporal dependence of returns in the predictive system using the Wold representation MA(∞) of the AR(1) process[10]. A similar representation can be done for the CIR-type process applying backward iteration of equation (6) and assuming the positivity of the trajectory, which yields:

$$\mu_t = E_r + \sum_{i=0}^{\infty} \beta^i w_{t-i} \sqrt{\mu_{t-1-i}}. \tag{8}$$

Equation (8) shows that, unlike the AR system, in our setting past innovations on expected returns occurring at times when the level of μ_t is relatively higher, have a higher impact in future expected returns than innovations occurring at times of low μ_t, everything else being equal.

As discussed in section 2.5, this dependence on the level of μ_t implies variations in both, the conditional variance of realized returns and the conditional variance of expected returns, and such variance variations are correlated with variations in the level of expected returns. If the change effect prevails over the level effect (as expected), these model predictions are in line with empirically established properties of the variance of returns.

[9]Similar processes such as Constant Elasticity of Variance (CEV), which we also consider on the derivation of the Kalman Filter in Appendix C, present the same type of mechanics.

[10]For the Wold representation of an AR(1) process see for instance Engle and Granger (1987), Anderson (2011), Cochrane (2005) chapter 6.

These effects are absent in the AR system, in which the filtered conditional variances of realized and expected returns are constant over time after a small number of steps in the Kalman filtering procedure.

Using equations (1) and (8), the return k periods ahead can be written as:

$$r_{t+k} = E_r + \sum_{i=0}^{\infty} \beta^i w_{t+k-1-i} \sqrt{\mu_{t+k-2-i}} + u_{t+k}. \tag{9}$$

Using equation (9) it can be shown that the autocovariance of r_t is equal to (cf. Appendix B for the derivation):

$$\mathrm{Cov}(r_{t+k}, r_t) = \beta^{k-1} \left(\underbrace{\beta \sigma_\mu^2}_{\text{level effect}} + \underbrace{\sigma_{uw} \mathbb{E}(\sqrt{\mu_{t-1}})}_{\text{change effect}} \right), \tag{10}$$

where σ_μ^2 is the unconditional variance of μ and is given by:

$$\sigma_\mu^2 = \frac{\sigma_w^2 E_r}{1 - \beta^2}. \tag{11}$$

In Pástor and Stambaugh (2009)'s predictive system the expressions for the autocovariance of r_t and the variance of μ are equal to equations (10) and (11), except that the unconditional expectation terms, $\mathbb{E}(\sqrt{\mu_{t-1}})$ in (10) and E_r in (11), do not appear. As we will see, equations (10) and (11) have important implications for the predictive system.

The R^2 of the regression of r_{t+1} on μ_t captures the fraction of variance in r_{t+1} explained by variations in μ_t. Hence it measures the level of predictability of r and is defined as,

$$R^2 = \frac{\sigma_\mu^2}{\sigma_r^2} = 1 - \frac{\sigma_u^2}{\sigma_r^2}, \tag{12}$$

then, a lower variance for μ implies a lower R^2. From the corresponding expressions for σ_μ for the AR and CIR systems, we have,

$$R_{ar}^2 = \frac{\sigma_{w,ar}^2}{\sigma_r^2 (1 - \beta_{ar}^2)}, \tag{13}$$

$$R_{cir}^2 = \frac{\sigma_{w,cir}^2 E_r}{\sigma_r^2 (1 - \beta_{cir}^2)}, \tag{14}$$

where the ar and cir subscripts indicate the system considered. In what follows we discuss two opposite cases to analyze the theoretical implications of the system modification. First, we keep σ_w, β and ρ_{uw} equal in both systems, which yields differences on R^2 and on the relative importance of the change and level effects. Second, we assume that both models have the same R^2, β and ρ_{uw}, in which case the change and level effects have the same relative weights in both systems and the parameter adjustment to get the same R^2 is done trough σ_w. Afterwards in sections 2.4 and 2.5 we further pursue these two comparative analyses with numerical examples.

On the one hand, notice that equations (10) and (11) imply that, if $\mathbb{E}(\sqrt{\mu_{t-1}}) < 1$ and $E_r < 1$ (a very mild assumption given historical returns), in the CIR system the variance of expected returns and the autocovariance of returns would be lower than in the AR system, everything else equal, that is, assuming σ_w, β, ρ_{uw} are equal for both models. Intuitively, this result steams from the fact in the CIR system μ_t must vary in a limited space of feasible values, unlike in the AR system for which μ_t can take negative values. Mechanically, we can see that expected returns innovations w are weighted by the previous value of $\sqrt{\mu}$ in equation (7) and thus, the size of the random shock in μ_{t+1} is much smaller in the CIR system than in the AR system when μ_t approaches zero, everything else being equal.

As Pástor and Stambaugh (2009) point out, the level and change effects can be "mapped" into the autocovariance of r. For $\rho_{uw} < 0$, whenever $\beta\sigma_\mu^2 < -\sigma_{uw}\mathbb{E}(\sqrt{\mu_{t-1}})$ returns are negatively autocorrelated, and the change effect prevails. Hence, ρ_{uw} needs to be "sufficiently negative" for this to happen. Setting σ_w, β, ρ_{uw} in the CIR system equal to the corresponding parameters of the AR system induces a change in the relative weight of the two terms reflecting the change and level effects in equation (10) with respect to the AR system, through the $\mathbb{E}(\sqrt{\mu_{t-1}})$ and E_r terms in the variance and autocovariance expressions. Assuming $\mathbb{E}(\sqrt{\mu_{t-1}}) \approx \sqrt{E_r}$, and $E_r < 1$ then $E_r < \sqrt{E_r}$; thus the change effect would have a larger relative weight in the return autocovariance of the CIR system (10) relative to AR system, keeping σ_w, ρ_{uw} and β constant (regardless of the level of E_r as long as $E_r < 1$).

Indeed, in the CIR system, the knife-edge value of ρ_{uw}, i.e., the value such that the change and level effects exactly offset each other in (10) is:

$$\text{k-e } \rho_{uw} = \frac{-\beta_{cir}\ \sigma_{w,cir}}{\sigma_u(1-\beta_{cir}^2)} \times \frac{E_r}{\mathbb{E}(\sqrt{\mu_{t-1}})}. \tag{15}$$

This expression is equal to the knife-edge of the AR system, except for the ratio $\frac{E_r}{\mathbb{E}(\sqrt{\mu_{t-1}})}$ (which is absent in the AR system). As a consequence, for equal σ_w, β, ρ_{uw}, the knife-edge value of ρ_{uw} for the CIR system would be closer to zero (i.e., less negative) than for the AR system as $\frac{E_r}{\mathbb{E}(\sqrt{\mu_{t-1}})} < 1$. This would imply a less restrictive condition on the level of ρ_{uw} for the change effect to prevail.

On the other hand, notice that the autocorrelation at k lags of returns in the CIR model can be written as a function of R^2, β and ρ_{uw},

$$\mathrm{Corr}(r_{t+k}, r_t) = \beta_{cir}^{k-1}\left(\beta_{cir}R_{cir}^2 + \rho_{uw}\sqrt{(1-R_{cir}^2)R_{cir}^2(1-\beta_{cir}^2)}\frac{\mathbb{E}(\sqrt{\mu_{t-1}})}{\sqrt{E_r}}\right), \tag{16}$$

which follows from equation (10) and noticing that $\sigma_\mu^2 = R^2\sigma_r^2$, $\sigma_u^2 = (1-R^2)\sigma_r^2$ and $\sigma_{w,cir}^2 = \frac{R_{cir}^2\sigma_r^2(1-\beta_{cir}^2)}{E_r}$. The autocorrelation expression (16) is equal to the autocorrelation of the AR system (see Pástor and Stambaugh, 2012, equation 22) except for the ratio $\frac{\mathbb{E}(\sqrt{\mu_{t-1}})}{\sqrt{E_r}}$. Assuming $\mathbb{E}(\sqrt{\mu_{t-1}}) \approx \sqrt{E_r}$, the term cancels out[11] in (16), thus if the two systems have equal R^2, β and ρ_{uw} then the level and change effects have the same relative weight in both systems. Another way to see this, is equating (13) and (14), i.e., if $R_{ar}^2 = R_{cir}^2$ then $\sigma_{w,ar}^2 = \sigma_{w,cir}^2 E_r$ and the extra ratio in the knife-edge formula (15) disappears, yielding the same expression in both systems.

Equation (14) shows that R_{ar}^2 and R_{cir}^2 are likely to diverge whenever E_r is relatively close to zero, while they should tend to be closer as E_r is further from zero (more on this in section 3.1). Thus, the difference in R^2 between the two systems is likely to be more accentuated for higher return frequencies (e.g., E_r is closer to zero for quarterly than for yearly return).

Investors may have priors on R^2 (as in Pástor and Stambaugh, 2009; Kvašňáková, 2013) or β (as in Pástor and Stambaugh, 2009), while intuition for σ_w^2 is less clear and σ_r^2 is almost observable and strictly equal for both models. Thus, in our out-of-sample Bayesian analysis (section 3.3) we use the same priors as in Pástor and Stambaugh (2009) for β and ρ_{uw} in both systems, and choose prior distributions for $\sigma_{w,ar}^2$ and $\sigma_{w,cir}^2$ such that the prior distribution of R^2 is equal in both systems. As we will see from the posterior distributions of β (Panels A and B of Figure 10), β_{cir} tends to be smaller than β_{ar}, and to steadily decrease over time from 1975 to 2012. Hence, μ_t tends to be less persistent in the CIR system. This result is consistent with the discussion of equation (7) above, since μ should have a higher speed of mean reversion (lower β) whenever it approaches zero; an effect absent in the AR system. The result indicates that, while most of the adjustment in parameters to obtain the same R^2 happens through σ_w, some of it can be due to differences in β. Notice that the relative weight of the term reflecting the level

[11]The square root of the sample average (unbiased estimator of the expected value) is a consistent although biased estimator of the expected value of the square root of the sample average (see for instance Barreto and Howland, 2006, p. 396). In our in-sample analysis of sections 2.4 and 2.5, we find empirically that the ratio $\frac{\mathbb{E}(\sqrt{\mu_{t-1}})}{\sqrt{E_r}}$ ranges between 0.99 and 1.00 for the R^2 and ρ_{uw} considered.

effect in equation (10) increases with the level of β. Thus, the more persistent expected returns are, the more important the level effect is. Conversely, the lower the variance of expected returns, σ_μ^2, the weaker the level effect relative to the change effect, indicating that the change effect can be even more likely to prevail in the CIR system.

Another variation of the widely used AR(1) representation for expected returns has been introduced in Van Binsbergen and Koijen (2011). In their specification, when μ_t gets closer to zero, its value is more likely to be pushed back to its long term mean rather than decrease, as in the CIR system. In Van Binsbergen and Koijen (2011), this effect is due to the time-varying autoregressive parameter but the conditional heteroscedasticity of μ introduced in our predictive system is not present in their model.

In order to further explore the implications of the model, in what follows we perform an in-sample comparison of the two systems for the same stock market series. In section II.B, Pástor and Stambaugh (2009) tested different values for ρ_{uw} to see its impact on the predictive system outputs. Their analysis is performed assuming $\beta = 0.9$, $R^2 = 5\%$ and using the sample estimates for σ_r and E_r in the system without predictors. In that case, all other parameters needed to estimate μ are functions of the parameters mentioned. The last row of Table 1 presents the numerical values of the parameters used in Pástor and Stambaugh (2009) section II.B, adjusted using the updated sample estimates for E_r and σ_r. Hereafter, we explore in section 2.4 the implications for the CIR system by first setting $\sigma_{w,cir} = \sigma_{w,ar}$ (leading to different R^2 for each system), which corresponds to the first row in Table 1 for the CIR system and the last row for the AR system, and then in section 2.5 comparing the two systems with the same level of R^2 (using the parameters in the last row of Table 1 for both systems).

2.4 In-sample analysis of AR and CIR systems with implied differences in R^2

In order to explore the role of ρ_{uw} in determining expected returns in the predictive system, Pástor and Stambaugh (2009), section II.B, work with a base-case parameter set for β, σ_w, σ_r, E_r, and R^2 in the AR system. In what follows, we present a similar steady-state in-sample analysis for the CIR model and compare our results with the AR system for the same values of β, σ_w, σ_r and E_r for both systems. The parameters in the first row of Table 1 correspond to the CIR system while the last row corresponds to the AR system of Pástor and Stambaugh (2009) section II.B (for the updated sample).

As pointed out above, using the CIR-based system we expect a lower variance for μ (and thus a lower R^2: 5% and 0.09% for the AR and CIR respectively) for the same σ_r^2 and σ_w^2. Hence, for given values of σ_r^2 and σ_w^2, this implies a greater variance for u with respect to the AR system. Keeping the same variance of realized returns, we use the same value of 0.0082 for σ_w, which implies $\sigma_u = \sqrt{\sigma_r^2 - \sigma_\mu^2} = \sqrt{\sigma_r^2 - \frac{\sigma_w^2 Er}{1-\beta^2}}$, leading to a slightly higher value of $\sigma_u = 0.0838$, compared to 0.0817 in the AR system.

For these parameter values and assuming $\frac{E_r}{\mathbb{E}(\sqrt{\mu_{t-1}})} \approx \frac{0.0177}{\sqrt{0.0177}} = 0.133$ yields a knife-edge value (15) in the CIR system of $\rho_{uw} \approx -0.06$. This contrasts with the knife-edge value of $\rho_{uw} = -0.47$ of the AR system[12]. This implies that the change effect can prevail over the level effect in our setting for a much larger range of values of ρ_{uw} with respect to the original system, everything else equal. In other words, there is a less restrictive condition on the value of ρ_{uw} for the change effect to dominate. This suggests that in the CIR version of the predictive system lagged returns deviations from E_r are more likely to have a negative weight on conditional expected return estimates. In this section, we perform our analysis using three values of ρ_{uw} in each system, i.e., -0.85 (this value corresponds to a dominant change effect for both systems and is used in the base-case analysis of Pástor and Stambaugh, 2009), the midpoint between -0.85 and the knife-edge values corresponding to each level of R^2 and the k-e ρ_{uw} (equivalent to historical mean, i.e., nil predictability benchmark)[13]. Hence as explained above, the three values of ρ_{uw} considered for the AR system will be much closer to each other than the values of ρ_{uw} for the CIR system, due to the larger range of ρ_{uw} corresponding to a dominant change effect in our modified system. It is interesting to note that the knife-edge value for the AR system (-0.47) is almost equal to the second ρ_{uw} considered for the CIR system (-0.46), i.e., for ρ_{uw} around -0.46 the change effect is nil for the AR system whereas it is dominant for the CIR system, in this configuration.

Using the Kalman filter, the conditional expected return can be written as the unconditional expected return mean plus linear combinations of past return forecast errors[14], where forecast error for the return in each period is defined as $\epsilon_t = r_t - \mathbb{E}(\mu_t|D_{t-1})$. Then

[12]In any case, μ_t being a priori less than one (which is a reasonable assumption), we should have $\frac{\mathbb{E}(\mu_{t+1})}{\mathbb{E}(\sqrt{\mu_t})} = \frac{E_r}{\mathbb{E}(\sqrt{\mu_t})} < 1$ which is in line with our analysis.

[13]We dot not consider a value of ρ_{uw} corresponding to a dominant level effect as there are several academic results in favor of a dominant change effect (see section 2.1).

[14]In the case where predictors are used, an additional term containing innovations in the predictors is added to equations (17), (18) and (20), for details, see Appendix C.

the expected return conditional on the history of returns can be written as

$$\mathbb{E}(\mu_t|D_t) = E_r + \sum_{s=0}^{\infty} \lambda_s \epsilon_{t-s}, \tag{17}$$

where $\lambda_s = \beta^s m$, and m is the steady-state filter parameter. The conditional expected return can also be written as a function of past returns instead of past forecast errors as follows:

$$\mathbb{E}(\mu_t|D_t) = E_r + \sum_{s=0}^{\infty} \omega_s (r_{t-s} - E_r), \tag{18}$$

where, in steady state,

$$\omega_s = (\beta - m)^s m. \tag{19}$$

Equations (17) and (18) have the same structure than the equivalent expression in the original predictive system, but the coefficient m is modified, leading to different predictions. The derivation of the steady-sate values of m for the CIR system is provided in the Appendix C, equation (63).

The conditional expected return depends on the true unconditional mean, E_r. Using the sample mean to estimate E_r in equation (18) and truncating the summation in the right-hand side to $s = t - 1$, yields an estimate of the expected return $\mathbb{E}(\mu_t|D_t)$ equal to a weighted average of past returns, i.e.,

$$\mathbb{E}(\mu_t|D_t) = \sum_{s=0}^{t-1} \kappa_s r_{t-s}, \tag{20}$$

where

$$\kappa_s = \frac{1}{t}\left(1 - \sum_{l=0}^{t-1} \omega_l\right) + \omega_s, \tag{21}$$

and $\sum_{s=0}^{t-1} \kappa_s = 1$. This expression has the same form than in the AR system. However, the ω_s are functions of m, which has a different expression in the CIR system that depend on the values of μ. The expression of m for the CIR system is provided in equation (63) of Appendix C.

Similar to the AR(1) predictive system of Pástor and Stambaugh (2009), whenever the change effect is greater than the level effect, i.e., the covariance term in the autocovariance (10) is larger than $\beta\sigma_\mu^2$, then $m < 0$ for sufficiently negative ρ_{uw} (see Appendix C). However, for given values of β, σ_w and σ_r, in the CIR-based version of the system the (absolute) values of m are much smaller in magnitude than the ones implied by the

AR-based original system. This also implies smaller values (in absolute value) of κ_s (presented in Figure 1) and λ_s, ω_s (which look very similar to κ_s and are unreported for space consideration).

Figure 1 displays the respective values of κ_s (weights on lagged returns), for the AR-based predictive system (as given in Figure 1, Panel B of Pástor and Stambaugh, 2009), and for the CIR-based predictive system, for the base-case parameter set aforementioned. First, as the figure shows, the range for κ_s in the CIR-based system is much closer to zero for different values of ρ_{uw} than for the original AR(1)-based system, with $\kappa_s \in [-0.0062, 0.0046]$ for the CIR and $\kappa_s \in [-0.0387, 0.0076]$ for the AR system. This implies that large lagged returns have a much smaller impact in future estimates of expected returns (see equation 20) in the CIR system, which is consistent with the lower variance of μ implied by equation (11). Second, for dominant change effect, while the AR system in this case gives negative weights to past returns of up to 40 lags (10 years) and positive weights to older returns, the CIR system assigns negative weights to past returns only up to 10 lags (2.5 years) and positive weights otherwise. This can be interpreted as the CIR system having a change effect with shorter memory. Third, the past returns weights in the CIR system is much closer to an equal weighted average of past returns (i.e., historical mean) while the AR system presents more disperse weights. Fourth, the weights of the CIR system are less sensitive to the level of ρ_{uw} than the weights in the AR system, due to the lower R^2.

Interestingly, from Figure 1 it can also be observed that the rate of decay of κ_s is much larger for the CIR system than for the AR system. As a consequence, the impact of the latest return, r_{t-1} is much larger than the impact of the return from two periods ago, i.e., $|\kappa_1| \gg |\kappa_2|$ in the CIR system than for the AR system. This implies for instance that, although variations in the AR are larger in absolute terms, a negative return after a series of positive returns will generate a larger correction, *in relative terms*, of $E(\mu_t|D_t)$ in the CIR system.

Panel A of Figure 2 reproduces the time series of the quarterly equity excess return estimates from Pástor and Stambaugh's AR system for the three different values of ρ_{uw} considered (similar to Figure 2 of Pástor and Stambaugh, 2009). Panel B of Figure 2 presents the conditional equity premium estimates using the CIR system with equal σ_w, β, E_r and σ_r (which correspond to the parameters in the first and last rows of Table 1 for the CIR and AR systems respectively). A striking result is the large difference in variance between the two series. While the AR-based quarterly return estimates vary between -0.2% (-0.8% annually) and +5.2% (+20.8% annually) for $\rho_{uw} = -0.85$, for the

CIR system, estimates of μ are well above zero and vary between +1.4% (+5.7% annually) and +2.6% (+10.2% annually). Besides this, it is noteworthy that the estimates from the AR system are much more sensitive to changes in ρ_{uw}.

Recall the analysis above is done for a given level of σ_w in both models. Hereafter we present a different analysis that explores the model implication by adjusting σ_w for the CIR system in order to keep R^2 equal in both models.

2.5 Can the CIR system explain the countercyclical predictability of the dividend-price ratio?

In the previous subsection, keeping the same value of σ_w yields a different value of R^2 for each system, due to the different dynamics of μ_t. The R^2 values were 5% and 0.09% for the AR and CIR systems respectively. We now set $R^2 = 5\%$ in both systems, maintain β, σ_r and E_r equal and set $\sigma_{w,ar}$ and $\sigma_{w,cir}$ such that the variance of μ is equal to 5% of the returns sample variance, i.e., $\sigma_\mu^2 = 0.05 \times \sigma_r^2$. The resulting parameters of both systems are reported in the last row of Table 1. As mentioned above, if R^2 and β are equal in both models, then the knife-edge value of ρ_{uw} is equal in both systems as well (using the approximation $\mathbb{E}(\sqrt{\mu_{t-1}}) \approx \sqrt{E_r}$).

For a given level of R^2 and β in both systems (adjusting the level of σ_w so as to obtain the same R^2) we find that the two systems produce very similar numerical results for μ if the steady-state formula for m is used. However, while the finite-sample estimates of m_t converge fast to the steady-state value in the AR system, this is not the case for the CIR system. In effect, given its dependence on the current level of μ, the finite-sample values of m_t present significant variations over time around the steady-state value. This feature has important implications for the conditional variance of μ which also implies a time-varying predictability of returns.

Whenever the change effect prevails, i.e., ρ_{uw} is sufficiently negative, during long stock market declines the level of μ increases which induces an increase in its conditional variance. This link between the level and variance of μ can be seen directly from the modified dynamics of μ with the CIR system in equation (6), where the standard deviation factor is $\sqrt{\mu}\sigma_{w,cir}$ which contrasts with the constant $\sigma_{w,ar}$ of the AR system. Figure 3 Panel A presents the square root of the filtered values of $\text{Var}(\mu_t|D_t)$, for the CIR and AR systems (without predictors) in finite-sample, when ρ_{uw} is set to -0.85. Moreover, Panel A of Figure 3 also presents the conditional variance of fitted values from a standard

OLS predictive regression using a rolling window with the latest 30 years of data as in Fama and French (1989), using the dividend-price ratio as the predictor variable x. Hence for the latter regression, we calculate the conditional variance of μ as: squared slope b^2, multiplied by the variance of x (i.e. the dividend-price ratio), both estimated using the latest 30 years of data. We observe that the conditional variance of μ for the CIR system increases significantly during economics recessions (grayed areas in the figure) as reported by the National Bureau of Economics Research. Similar increases are observed for the predictive regression using the dividend-price ratio, while the conditional variance of μ in the AR system is constant over time. A potential consequence of this is that the portion of variance of realized returns explained by variations in μ, i.e., R^2, might increase during market downturns.

Henkel et al. (2011) finds that the conditional R^2 of a predictive Regime-Switching Vector Autoregressive model presents significant variations across the business cycle, using several predictors including the dividend yield. In order to compute a time-conditional R^2 in our rolling predictive regression we calculate,

$$R^2(t+1|D_t) = \frac{\sigma^2_\mu(t|D_t)}{\sigma^2_r(t+1|D_t)}, \tag{22}$$

where $\sigma^2_\mu(t|D_t)$, the conditional variance of μ, is calculated as explained in the previous paragraph, and $\sigma^2_r(t+1|D_t)$, the conditional variance of returns, is estimated using the sample estimate for the variance of r with the latest 30 years of data. Notice that the term (22) is similar in spirit to to the conditional R^2 considered in Henkel et al. (2011).

Similarly, we calculate the conditional R^2 of the CIR and AR predictive systems as given by the right-hand side of (22), using the corresponding filtered values for $\sigma^2_\mu(t|D_t)$ and $\sigma^2_r(t+1|D_t)$ which correspond to the quantities Q_t and the first component of S_{t+1} in the notation of the Kalman Filter in Appendix C. Panel B of Figure 3 presents the conditional R^2 of equation (22), i.e., the ratio of the conditional variances of μ_t and r_{t+1} for both systems and for the (rolling) predictive regression using the dividend yield.

The time series of the conditional R^2 of the CIR system without predictors and the corresponding series of the predictive regression present remarkably similar dynamics, increasing and decreasing over time in tandem. Moreover, the conditional R^2 increases during economic recessions. We find a qualitatively equivalent result when using the same conditional variance estimate of returns used in the predictive regression instead of the first component of S_{t+1} in the formula of the conditional R^2 of the CIR system (unreported for space considerations). This result confirms the counter-cyclicality of the predictability of the dividend yield documented by Henkel et al. (2011). More importantly, it shows that the CIR predictive system *without predictors* can structurally explain the countercyclical dynamics of return predictability, an effect absent in the original AR system which presents a constant conditional R^2 as shown in Figure 3 Panel B. Interestingly, in both panels of Figure 3, we note that after 2000, the curves corresponding to the CIR system and dividend-yield regression start to diverge significantly in terms of level. Goyal and Welch (2003) highlight an instability of dividend price ratio autoregression coefficient, which has increased from about 0.4 in 1945 to about 0.9 in 2000 according to their estimation procedure. This observation may explain the behavior of the predictive regression curves in Figure 3 as an increasing autoregression coefficient implies a lower conditional variance in dividend yield and thus a lower conditional R^2. This effect is absent in the CIR system as the parameters used in this configuration are constant over time.

Another interesting implication of the modified dynamics of μ is that the finite-sample CIR system produces a conditional estimate of the variance of realized returns that varies over time as well. On the other hand, the conditional variance of r_t for the AR system converges fast and stabilizes to a constant steady-state value. Figure 4 presents the evolution of the finite-sample filtered volatility of r_t, i.e., the square root of $\mathrm{Var}(r_t|D_{t-1})$ for both systems (given in equation 45 in Appendix C), with $\rho_{uw} = -0.85$ (dominant change effect for both systems), as well as the log-cumulated excess returns of the market. In contrast with the AR system, for which the filtered volatility reaches fairly fast a steady value (this behavior is observed for all ρ_{uw}, unreported), in the CIR system the volatility evolves depending on the level of μ, and therefore on the level of r. We can see on Figure 4 that the volatility reaches its maximum (resp. minimum) values when μ_t is high (cf. dotted lines on Figure 2), thus during periods of low (resp. high) realized returns. In other words, the conditional volatility of r in the CIR model is positively correlated with μ_t. At the same time, μ is negatively correlated with r, and hence the variance of returns implied by the finite-sample filter of the CIR system is countercyclical, as illustrates Figure 4. This model prediction is consistent with the commonly observed fact that the volatility of stocks increases as its price falls, for standard empirical variance estimates (see for instance French, Schwert, and Stambaugh, 1987; Engle, 2001; Ait-Sahalia and Kimmel, 2007).

The changing dynamics of μ in the CIR system can also be mapped in the dynamics of the coefficients connecting the realized returns and expected return estimates of the system. As Figure 5 shows, unlike the AR system, the CIR system has a "conditional-memory" (conditional on the phase of the market). In effect, the lagged coefficients κ and also λ, ω (unreported for space consideration) of the CIR system are different, whether they are calculated after a series of positive realized returns or after a series of negative realized returns. In order to illustrate the "conditional-memory" of the CIR system we consider two sub-periods of the total sample, i.e., 1952Q1 to 1999Q4 (end of the bull market of the IT bubble), and 1952Q1 to 2002Q4 (end of the bear market after the IT bubble bursts). Figure 5 presents the corresponding values of the finite-sample coefficients κ (the behavior of λ and ω coefficients is similar), for both systems calculated at the end of the two sub-periods. From the figure we see that in the CIR system the latest observations have a larger impact in expected returns following a falling market than after a rising market (when change effect dominates). This effect is induced by the proximity of μ_t to its lower bound (zero) after a long bull market, as further positive returns cannot push μ_t much further down.

We observe a remarkable difference in the shape of the corresponding coefficients curves for the two systems. Indeed, the rise from the most negative values of the coefficients corresponding to the lower lags (i.e., most recent values) toward zero of the curve of κ, is much sharper in the CIR system. As a consequence, similar to the former analysis with equal σ_w, this implies that the latest return will induce a larger (relative) correction of the estimate of μ_t in the CIR system relative to the AR system. To see this, consider any two lag indices l_1 and l_2 such that $0 < l_1 < l_2 < 20$ (notice that the lower lag l_1 refers to a more recent observation than l_2) in Figure 5, and remark that[15] $\frac{|\kappa^{cir}_{l_1}|}{|\kappa^{cir}_{l_2}|} > \frac{|\kappa^{ar}_{l_1}|}{|\kappa^{ar}_{l_2}|}$. Unlike the former analysis of section 2.4, in which κ (and λ, ω) corresponding to the latest observation were much smaller in absolute value in the CIR relative than in the AR system, in this case, the coefficients of the lower lags are also larger in absolute terms for the CIR system, i.e., $|\kappa^{cir}_1| > |\kappa^{ar}_1|$. Thus, the CIR system gives more (negative) weight to the latest observations than the AR system. This effect is due to a higher time dependence of m_t with respect to the corresponding terms in the AR system (cf. Appendix C), and it is also observed in the weights calculated at the end of the total sample, 2012Q4 (unreported for space considerations).

Panel A and C of Figure 2 present the filtered equity premium, $\mathbb{E}(r_{t+1}|D_t)$, for the two systems on the overall sample period when their R^2 is set to 5%. We note in particular that the level of μ reaches larger maximum values in the CIR system (Panel C) than in the AR system (Panel A), while it tends to slow its variations much faster as it approaches zero as suggested by the dynamics of the CIR model.

[15]We also observe $\frac{|\lambda^{cir}_{l_1}|}{|\lambda^{cir}_{l_2}|} > \frac{|\lambda^{ar}_{l_1}|}{|\lambda^{ar}_{l_2}|}$ and $\frac{|\omega^{cir}_{l_1}|}{|\omega^{cir}_{l_2}|} > \frac{|\omega^{ar}_{l_1}|}{|\omega^{ar}_{l_2}|}$, unreported here.

3 Empirical out-of-sample analysis

In this section we present an out-of-sample analysis of return predictability using a Bayesian approach for estimating the predictive systems. First, we discuss the implications of using the CIR system on the priors on parameters distributions. Then we conduct an exploratory analysis using point estimate parameters for implementing the systems. Finally, we present our Bayesian analysis and out-of-sample results.

3.1 What are plausible values for R^2, β and E_r?

The empirical implementation and estimation of the predictive system in a Bayesian setting needs a set of priors on the distribution of the parameters involved. These priors should represent a plausible parameter set, compatible with the hypotheses behind the system. Unlike Pástor and Stambaugh (2009)'s AR system, one of the motivations to use the CIR system is the assumption that the expected return process is unlikely to be negative. Hence, the priors used to implement the CIR system should be compatible with this hypothesis. In this section we discuss the plausible values for the R^2 of the regression of r_{t+1} on μ_t, the persistence parameter β and the long-term mean E_r using the same market and sample as in Pástor and Stambaugh (2009) and in sections 2.4 and 2.5 above.

The CIR continuous time model ensures the non-negativity of the process μ in equation (5) if the parameters respect the condition $\kappa\theta \geq \frac{\sigma^2}{2}$ (Feller, 1951), which can be translated in terms of the parameters of the discretized process as :

$$(1-\beta)E_r \geq \frac{\sigma_w^2}{2}. \tag{23}$$

By definition of σ_μ, condition (23) also implies an upper bound for σ_μ^2 in the CIR model:

$$\sigma_\mu^2 = \frac{\sigma_w^2 E_r}{1-\beta^2} \leq \frac{2(1-\beta)E_r^2}{1-\beta^2}. \tag{24}$$

Furthermore, using an estimate for the variance of realized returns σ_r^2, this condition also provides an upper bound for the R^2 from the regression of r_{t+1} on μ_t since by definition (12) it follows that,

$$R^2 = \frac{\sigma_\mu^2}{\sigma_r^2} \leq \frac{2(1-\beta)E_r^2}{(1-\beta^2)\sigma_r^2}. \tag{25}$$

This means that, for a given set of plausible values for E_r and β, the CIR positivity condition (23) restrains the possible value set for σ_μ, and for R^2 for a given estimate of σ_r. On the other hand, there is no such internal coherence restriction in the AR system as there is no positivity constraint. From condition (23) and equation (25), it follows that to estimate the upper bound of R^2 for the CIR system, we need to get: 1) E_r^*: the highest plausible value of E_r, and 2) β^*: the lowest value for β. Considering the return sample used in Pástor and Stambaugh (2009), to estimate the highest plausible value of E_r, we calculate $\max(\hat{E}_r(1,..s_0))$ for a sample of size s_0 for $s_0 = \{80, 81, ...208\}$, where 80 points corresponds to a minimum sample period of 20 years and 208 is the total sample size of quarterly data from 1952-Q1 to 2003-Q4 used in Pástor and Stambaugh (2009), which yields $E_r^* = \max(\hat{E}_r(1,..s_0)) = 0.0226$ (9% p.a.). Second, we use the 5% quantile of the prior distribution of β in Pástor and Stambaugh (2009), i.e., $\beta \sim \mathcal{N}(0.99, 0.15^2)$, hence $\beta^* = 0.99 + 0.15 \times z_{(5\%)} = 0.743$, where $z_{(5\%)}$ denotes the 5% quantile of a random variable with standard normal distribution. These two estimates for E_r^* and β^* within condition (23) together with the sample estimate of $\hat{\sigma}_r = 0.0837$ yields an upper bound for the R^2 in the CIR system of 8.37%. This upper bound is in line with Pástor and Stambaugh (2009)'s statement that a plausible prior distribution of R^2 would have most of its mass being below 5% and a mode around 1% for the US stock market quarterly data sample considered.

Note that, using the knife-edge formula of ρ_{uw}, it is possible to derive bounds for the persistence parameter β in order to have a plausible knife-edge value, for both, the AR and CIR systems for a given value of R^2. Considering a knife-edge value within the interval [-1, 0] is equivalent to not excluding the possibility that the change effect dominates over the level effect. Notice that, if $\rho_{uw} > 0$ there is no change effect at all. If the knife-edge value of ρ_{uw} of the AR system is between -1 and 0 then,

$$|\text{k-e } \rho_{uw}| = \left| \frac{-\beta\sigma_w}{\sigma_u(1-\beta^2)} \right| < 1. \tag{26}$$

From the definition of R^2 and σ_μ for the AR system, notice that, $\sigma_w^2 = (1-\beta^2)R^2\sigma_r^2$ and $\sigma_u^2 = (1-R^2)\sigma_r^2$. Replacing these in inequality (26) and squaring yield,

$$\frac{R^2\beta^2}{(1-R^2)(1-\beta^2)} < 1$$

$$|\beta| < \sqrt{\frac{C}{1+C}} \tag{27}$$

where $C = \frac{(1-R^2)}{R^2}$, which simplifies to $|\beta| < \sqrt{1-R^2}$. For instance, if $R^2 = 5\%$ the upper bound of β is 0.97. A similar calculation using the knife-edge value for the CIR system (equation 15) yields the bounds for β in the CIR model which are given by (27) but with $C = \frac{(1-R^2)}{R^2 E_r}$. The consistency condition (27) implies an inverse relationship between

R^2 and the maximum feasible persistence parameter of μ. If one believes that μ is very persistent and we do not preclude the possibility that the change effect dominates over the level effect then the R^2 cannot be very high and viceversa.

The positivity condition (23) implies a long-term mean strictly positive, i.e., $E_r > 0$. Pástor and Stambaugh (2009)'s prior distribution for E_r is Gaussian with a "large" 1% standard deviation around its sample mean, denoted $\hat{E}_r$ (see p. 9 of the internet appendix in Pástor and Stambaugh, 2009). In effect, the 1% quantile of such distribution is a negative number, which is incompatible with the CIR model assumption, especially if we consider that the presumably positive process μ should vary around E_r. Indeed, assuming that expected returns are non-negative implies that the long-run average of μ, E_r should be "far enough" from zero. Thus, a plausible value for the variance of the prior distribution of E_r should be lower than 1% for the CIR system. For instance, given that the prior for the distribution of E_r is symmetric, one may assume that the distance between the sample mean $\hat{E}_r = 0.0185$ and its highest plausible value of $E_r^* = 0.0226$ (see calculation above), is the same distance between $\hat{E}_r$ and a low quantile of its distribution. We deduce σ_{E_r} as follows. Assume $\hat{E}_r + \sigma_{E_r} z_{(5\%)} = \hat{E}_r - (E_r^* - \hat{E}_r)$, hence

$$\sigma_{E_r} = \frac{(\hat{E}_r - E_r^*)}{z_{(5\%)}},$$

which is 0.25% for the sample estimates mentioned above, thus about four times smaller than the prior standard deviation of 1% used in Pástor and Stambaugh (2009) for the AR model. This choice ensures a positive 1% quantile for the prior distribution of E_r.

3.2 Out-of-sample return prediction using point estimate parameters

In what follows we present the results of an exploratory analysis which consists in an out-of-sample return prediction exercise of quarterly returns of the value-weighted CRSP US aggregate stock market index in excess of the quarterly return on 1-month T-bills obtained from the Center for Research in Security Prices (CRSP). Following Pástor and Stambaugh (2009) and other studies we begin our sample in 1952, as in our in-sample analysis of sections 2.4 and 2.5. In order to address concerns regarding the dependence of predictability evidence on the oil price shocks period 1973−1974, we set our out-of-sample period to 1975-2012 for this first analysis. We also consider additional out-of-sample periods in the Bayesian analysis of section 3.3.

In order to estimate μ, we use the predictive systems without predictors, with fixed values for the models parameters. Every quarter we estimate E_r and σ_r using the prevailing sample estimates at each point in time and we set the R^2 of the regression of r_{t+1} on μ_t equal in both models, with $R^2 = 0.5\%, 1\%, 2\%, 3\%, 4\%, 5\%$. We consider two plausible values for the persistence parameter of $\beta = 0.9$ and then $\beta = 0.8$ (kept constant throughout the entire period). Furthermore, we perform predictions using several values of ρ_{uw}, i.e., from -0.95 to 0.95 with a step of 0.05. All other parameter values in the systems follow from their corresponding definitions.

Following former studies such as Goyal and Welch (2008), we assess out-of-sample predictive power with the out-of-sample (OS) R^2_{OS} introduced by Campbell and Thompson (2008),

$$R^2_{OS} = 1 - \frac{MSE_{pred}}{MSE_{mean}}, \tag{28}$$

where MSE_{pred} is the mean squared error of the model predictions and MSE_{mean} is the mean squared error of using the prevailing return's historical average as estimate of expected return. This metric evaluates weather a given system produces more accurate predictions than the no-predictability random walk benchmark, i.e., the prevailing historical average.

The overall results are presented in Figure 6 for $\beta = 0.9$, and Figure 7 for $\beta = 0.8$. The two figures present the R^2_{OS} for each system, as a function of ρ_{uw}, when setting the R^2 of the regression of r_{t+1} on μ_t to a given value in $\{0.5\%, 1\%, 2\%, 3\%, 4\%, 5\%\}$. The grayed areas in both figures correspond to values of ρ_{uw} implying a dominant change effect.

First, we observe that the highest R^2_{OS} for each system are obtained with different priors on R^2 and β. The highest R^2_{OS} across all parameter combinations considered corresponds to the CIR system (3.39%), with $\beta = 0.8$ and $R^2 = 2\%$. On the other hand, the highest R^2_{OS} for the AR system is 3.00% and corresponds to $\beta = 0.9$, and $R^2 = 4\%$ (where $\rho_{uw} = -0.95$ in both cases). This result is consistent with the intuition that the CIR system is more suited for lower levels of R^2 (as μ would have a lower variance) and thus with a less a stringent condition on ρ_{uw} to yield a dominant change effect. Furthermore, it also suggests that the CIR system may imply an expected return process with lower persistence, a result confirmed by the posterior distribution of β of the Bayesian analysis of the next section (3.3).

Also, we note that for all values considered for the R^2 of the regression of r_{t+1} on μ_t, for both AR and CIR systems, the R^2_{OS} are positive when ρ_{uw} is below its knife-edge value (except the case $\beta = 0.8$, $R^2 = 4\%, 5\%$ and $\rho_{uw} = -0.95$ for the CIR system[16]). Table 2 presents a summary of the highest and lowest R^2_{OS} obtained for each system as well as the corresponding parameters set. For both systems, the highest R^2_{OS} are obtained with $\rho_{uw} = -0.95$ (3% and 3.39% respectively for the AR and CIR system), whereas the lowest R^2_{OS} are obtained for $\rho_{uw} = 0.95$ (-5.81% and -5.62% respectively for the AR and CIR system). This result indicates that Pástor and Stambaugh (2009)'s believe that the change effect should dominate, is also consistent with the CIR system.

It is also interesting to notice that in all configurations with $\rho_{uw} \geq 0$ (only level effect) the less negative R^2_{OS} are obtained with the CIR system for all priors of R^2 considered.

In the next section, we explore the performance of both systems, considering different out-of-sample periods and different priors on the parameters. The results presented thereafter confirm the suggestions above.

3.3 Out-of-sample return prediction using Bayesian parameters estimates

Using the same data as in section 3.2, we conduct now an out-of-sample analysis of both predictive systems using the Bayesian parameter estimation procedure of Pástor and Stambaugh (2009). This procedure allows incorporating parameter uncertainty and specifying less informative prior distributions. Posterior distributions for the parameters are obtained using Gibbs sampling (see for instance Kim and Nelson, 1999). Following Pástor and Stambaugh (2009), we estimate the predictive systems parameters by simulating 76000 posterior draws, dropping the first 1000 as a "burn-in" period and take every third draw from the rest to obtain 25000 posterior draws. The overall Markov Chain Monte Carlo (MCMC) procedure and the posterior distributions for the AR system are described in the internet appendix of Pástor and Stambaugh (2009). We refer the reader to Appendix D for further details on the Bayesian parameter estimation procedure and the posterior distributions in the CIR system.

The predictive systems parameters are re-estimated on the first available date of each year in the sample, while predictions (estimates of μ) are computed each quarter using the data available at each point in time (thus running the filter with the same parameters over the year).

[16]The poor values of the R^2_{OS} for the CIR system, for "large" R^2 ($3\%, 4\%, 5\%$) and $\rho_{uw} = -0.95$, are mainly due to the first prediction which is very high given the negative return of -26.79% in 1974-Q3.

For the AR system, the priors distributions used are identical (except for E_r, see below) to those described in section B.5 of the internet appendix of Pástor and Stambaugh (2009). We thus refer to their initial paper for more details and provide only a brief description of the distributions. The prior for β, plotted in Panel B of Figure 9, is chosen to capture the belief that μ is persistent, i.e., β is smaller than one but not by much[17] $\beta \sim \mathcal{N}(0.99, 0.15^2) \times \mathbb{I}_{\beta\in(0,1)}$. The prior on E_r is slightly modified in order to use the same prior as the CIR system[18]: $E_r \sim \mathcal{N}(\bar{r}, \sigma^2_{E_r})$, where $\bar{r}$ denotes the mean of the returns $\{r_t\}$ available at the date of estimation and σ_{E_r} is choosen as described in section 3.1. We consider three prior distributions for σ_w, plotted in Panel B, D and F of Figure 8. The submatrix $\Sigma_{11} = \begin{bmatrix} \sigma_u^2 & \sigma_{uw} \\ \sigma_{wu} & \sigma_w^2 \end{bmatrix}$ has, for each prior, an inverted Wishart distribution: $\Sigma_{11} \sim IW(T_0\hat{\Sigma}_{11,0}, T_0)$, where T_0 is equal to one fifth of the available return sample size. The prior mean $\mathbb{E}(\Sigma_{11})$ is set according to: 1) a prior value $\bar{R}^2$ for its diagonal elements, and 2) our priors on ρ_{uw} (see below) for the non-diagonal elements[19]. The three different priors on σ_w, corresponding to different priors on the R^2 (and the variance of μ), are obtained by setting $\mathbb{E}(\sigma^2_\mu)$ equal to a given percentage of the prevailing sample return variance $\hat{\sigma_r}^2$, i.e., $\mathbb{E}(\sigma^2_\mu) = \bar{R}^2\hat{\sigma_r}^2$ and $\mathbb{E}(\sigma^2_u) = (1-\bar{R}^2)\hat{\sigma_r}^2$ for $\bar{R}^2$ equal to 2.5% (less predictability prior), 5% (prior used in Pástor and Stambaugh, 2009, denoted hereafter benchmark prior) and 10% (more predictability prior). The corresponding prior distributions of the R^2 are presented respectively in Panel A, C and E of Figure 8. Moreover, we consider two priors on ρ_{uw} used in Pástor and Stambaugh (2009), which are presented in Panel A of Figure 9: noninformative (flat between -0.9 and 0.9) and more informative (most of the mass below -0.71).

The priors used for the CIR system are the same for ρ_{uw}, β and R^2. The latter implies a prior distribution with higher levels of σ_w as shown in Panels B, D and F of Figure 8 and explained in section 2.5. As described in detail in section 3.1, the prior distribution of E_r used here for both the CIR and AR systems, is slightly different from the prior used in Pástor and Stambaugh (2009); thought we use the same method to estimate the mean, we use a lower variance, in order to preclude negative value draws for the long-term mean

[17] $\mathbb{I}_{\beta\in(0,1)}$ denotes here the indicator function equal to 1 if $\beta \in (0,1)$ and 0 otherwise. In our case, this corresponds to retain only draws of β satisfying the condition.

[18] We also tried to use the initial prior for E_r used in Pástor and Stambaugh (2009) for the AR system, i.e., $E_r \sim \mathcal{N}(\bar{r}, 0.01^2)$, but the out-of-sample results were poorer, i.e., negative R^2_{OS} for the AR system.

[19] We refer to page 9 of the internet Appendix of Pástor and Stambaugh (2009) for a description of the prior draw procedure of the non-diagonal elements.

of expected returns E_r.

Again, we use the R^2_{OS}, described in section 3.2, to assess the out-of-sample predictability of the systems. The statistical significance of R^2_{OS} is assessed using the F_{MSE} statistic proposed by McCracken (2007), which tests for equal MSE of the historical mean and system's conditional forecasts. It is given by:

$$F_{MSE} = (T - s_0)\frac{(MSE_{mean} - MSE_{pred})}{MSE_{pred}}, \tag{29}$$

where T stands for the total size of the sample periods and s_0 for the initial calibration sample. In our tables we use $*$, $**$ and $***$ to indicate statistical significance at 10%, 5% and 1% levels respectively.

The results are presented in Table 3 for four out-of-sample periods. The first one is the longest and is the same period considered in section 3.2 with point estimates: 1975-2012. We also consider three additional out-of-sample periods of 25 years, whose starting dates are spaced by 5 years: 1975-2000, 1980-2005, 1985-2010.

There are four clear conclusions from the results, 1) the more informative prior on ρ_{uw} implying a dominant change effect improves out-of-sample return forecasts for both systems, 2) for all periods and predictability priors considered, using the more informative prior on ρ_{uw}, the CIR system yields better out-of-sample estimates than the AR system, 3) for both systems, the highest R^2_{OS} for each out-of-sample period are obtained using different priors on the R^2 of the regression of r_{t+1} on μ_t, 4) the CIR system using the more informative prior on ρ_{uw} yields significantly better out-of-sample predictions than the prevailing historical average for all out-of-sample periods considered, for at least one R^2 prior (the R^2_{OS} is significant in 10 out of 12 priors and sample periods combinations considered).

Indeed, a comparison of the R^2_{OS} obtained with the noninformative prior and the more informative prior on ρ_{uw} for each system suggests that ρ_{uw} is more likely to be negative. For instance, in the longest out-of-sample period 1975-2012 (first panel of Table 3), the results obtained with the less predictability prior (prior on R^2 leading to the best results for both systems on this period), for the AR system, using the more informative prior on ρ_{uw} instead of the noninformative, leads to an increase of the R^2_{OS} from -1.03% to 0.28%. Although the sign of the latter R^2_{OS} is positive, the AR system predictions do not outperform significantly the prevailing historical average in terms of MSE. Regarding the CIR system, the noninformative prior on ρ_{uw} yields an R^2_{OS} of -0.30%, less negative than the AR system in this configuration but implying prediction errors still greater than the historical mean. On the other hand, using the more informative prior leads to a positive

R^2_{OS} of 1.21%, significant at the 5% level. Consequently using the CIR system with the more informative prior on ρ_{uw} produces significantly more accurate predictions than the historical mean of returns, which is not the case when using the original AR system in this case. The conclusions regarding the benefits of using the more informative prior on ρ_{uw} are confirmed by the results obtained for each out-of-sample sub period considered, and for each prior on R^2.

We also observe on Table 3 that, the highest R^2_{OS} for the CIR system is systematically greater than that of the AR system (for all periods), confirming results of section 3.2: there exist benefits in terms of out-of-sample prediction in using the CIR system instead of the AR system. Hence, the new features of the CIR system (conditional heteroscedasticity and non-negativity of expected returns) are consistent with the dynamic of the true unobservable expected return process.

Moreover, for both systems, we note that the prior on R^2 leading to the highest R^2_{OS} is different for each out-of-sample period considered. In other words, assuming that return predictability is lower or higher implies better out-of-sample estimates, depending on the period. Thus result suggest that the predictability of returns is in fact time-varying, which is in line with findings in Rapach et al. (2010) and Henkel et al. (2011). This result also suggest that the outperformance of the CIR system with respect to the AR system in our Bayesian analysis may be explained by the fact that the modified system incorporates expected returns heteroscedasticity.

Additionally, for all out-of-sample periods considered, using the more informative prior on ρ_{uw} leads to positive R^2_{OS} for both models for at least two of the priors on R^2. However, the significant R^2_{OS} are obtained for all periods for the CIR system (with different predictability priors on R^2), but only for two out-of-sample periods for the AR system (1975-2000 and 1980-2005).

Figure 10 presents for the longest out-of-sample period (1975-2012), the evolution of the posterior mean of the parameters β and ρ_{uw} (re-estimated on the first available date of each year in the sample) using the less predictability prior[20] on R^2 and the more informative on ρ_{uw}. We observe first that the average levels of these parameters differ for the AR and the CIR systems, and second that the posterior means are stable throughout the out-of-sample period for the AR system whereas they vary over time for the CIR system. Hence, the latter observation suggests that the true values of parameters of the system may change over time.

[20]Conclusions are similar using the benchmark prior and more predictability priors on R^2 (unreported).

We notice on Panels A and B of Figure 10 that the posterior mean of β for the CIR system (ranging from about 0.8 to 0.4) is always lower than its equivalent in the AR system (which is stable around 0.9). This observation suggests a less persistent expected return process for the CIR system and confirms the results obtained using point estimate parameters in section 3.2. Moreover, unlike for the AR system, the posterior mean of β for the CIR system is obviously not constant over time: β is stable around 0.6 from 1975 to 1982, increases to 0.7 by 1986, and then declines smoothly to 0.4 by 2012. This decrease in the level of β explains the progressively lower R^2_{OS} observed in the latest two sub periods (1980-2005 and 1985-2010). To see this, notice that according to equation (14), the lower β, the lower the proportion of returns variance explained by expected returns.

Furthermore, Figure 12 presents the evolution of the sample autocorrelation of returns at lag 1, using the available returns sample at each point in time. The similarity with the Panel B of Figure 10 is striking. Indeed, the CIR system, through the parameter β captures the fact that returns autocorrelation is not constant over time. A decrease in the latter is related to a decrease in the persistence parameter of expected returns using the CIR system.

Another divergence between the systems is observed on Panels C and D of Figure 10, where the posterior mean of ρ_{uw} is more negative for the AR system: between -0.6 and -0.8, than for the CIR system: between -0.2 and -0.5. This observation also confirms the results obtained in section 3.2: the value of ρ_{uw} should be closer to zero if we adopt the CIR system as the predictive model. Interestingly, similarly to the parameter β, the posterior mean of ρ_{uw} is relatively constant over time for the AR system while it is time-varying for the CIR system. Indeed, for the latter, the variations of ρ_{uw} in Panel D are virtually the mirror image of the variations of β in Panel B (sample correlation of -0.86 between the posteriors of β and ρ_{uw}), meaning that, as β decreases, the correlation between expected and unexpected returns becomes less negative.

These conclusions are confirmed by Figure 11, which presents the posterior distributions of parameters ρ_{uw} and β, corresponding to the estimations for 1992 and twenty years later in 2012 (last estimation). While the posterior distributions for the AR system are substantially the same, there exist clear differences for the CIR system between the two dates of estimation. Posteriors for ρ_{uw} are shifted to zero (less negative) between 1992 and 2012, as well as posteriors for β (much less close to one). Indeed, for the CIR system, in 1992, most of the mass of the distribution is below zero for ρ_{uw}, and between 0.6 and 1 for β. This is no longer the case in 2012, when the distribution of ρ_{uw} is much more neutral (less clearly negative, although the negative part is still heavier than the positive one), and posterior values of β are spread from 0 to 1 with a mode around 0.5.

In summary, the evolution of the posterior distributions and posterior means of the CIR system parameters highlight two differences with the AR system. First, the shape of the posterior distribution of β actually changes over the period studied, and the persistence of expected return steadily decreases during the last thirty years of our sample. This behavior is consistent with the autocorrelation of realized returns (at lag 1) which presents a similar decreasing trend over the period. Second, the posterior distribution of ρ_{uw} also varies throughout the sample, and it gradually becomes less negative toward the end of the sample.

Conclusion

The predictive system introduced by Pástor and Stambaugh (2009) provides a rich environment to estimate and analyse expected returns estimates and their relationship with past realized returns. The original version of the system assumes an AR(1) dynamic for the expected excess return process, which allows negative values and implies a constant conditional variance.

However standard equilibrium models with risk-averse investors predict non-negative expected excess returns on the market, and several empirical studies such as Rapach et al. (2010), and Henkel et al. (2011) found that return predictability in the stock market is stronger during recessions. The latter finding, together with the fact the variance of realized returns also increases during recessions implies that the conditional variance of expected returns must increase in economic downturns.

This paper introduces a modified version of the predictive system in which the expected returns process are unlikely to be negative and present a time-varying conditional variance, without any additional parameters. We find that the modified system without predictors can explain the counter-cyclical variations in the predictive power of the dividend-price ratio documented by Rapach et al. (2010), and Henkel et al. (2011).

Additional theoretical and empirical implications of the modified system include: 1) a lower variance of expected returns, 2) an expected return process with lower persistence than in the original system, 3) a less negative correlation between expected and unexpected returns, and 4) a varying conditional variance of realized returns produced by the system that is likely to be negatively correlated with realized returns. The former two implications are consistent with well documented features of stock returns, the first is in line with the optimal shrinkage forecasting methodology in Connor (1997), while the second is consistent with implications of the present-value model in Van Binsbergen and Koijen (2011). Furthermore, in out-of-sample tests we find that the modified system can produce significantly better return predictions than the historical average, and using a Bayesian approach, better forecasts than the original system.

In our Bayesian parameter estimation of the modified system, we also find that the persistence of expected returns is not constant over time and has steadily decreased for thirty years, in tandem with the sample autocorrelation of realized returns. Also, in our system the correlation between expected and unexpected returns becomes less negative toward the end of the sample period. These findings suggest that interesting further extensions of the the system would be an expected return process with a persistence parameter explicitly varying over time, and a time-varying correlation between realized and expected returns.

Appendix

A Discretization of CIR process

The Cox et al. (1985) (CIR) model is defined by the following Stochastic Differential Equation (SDE):

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t, \quad X_0 \geq 0, \tag{30}$$

where κ, θ and σ are constants, and W is a standard Brownian motion. This SDE has a level dependent diffusion term ($\sigma\sqrt{X_t}$) implying a conditional heteroscedasticity for the process X. Furthermore, the model can rule out negative values for X. Indeed, given that κ is a Lipschitz constant for the drift term of (30), Feller's test (c.f. Feller, 1951) for univariate stochastic process ensures that the following condition $\kappa\theta \geq \frac{\sigma^2}{2}$ implies $\mathbb{P}(\tau_0^x = \infty) = 1$, where $\tau_0^x = \inf\{t \geq 0 : X_t = 0\}$ and x refers to the case $X_0 = x \geq 0$. Details are given in Berkaoui, Bossy, Diop, et al. (2008).

The CIR equation (30) does not have an explicit closed-form solution for the process X, which is inconvenient to develop the economic intuition of the model. A direct Euler discretization of (30) is:

$$\mu_{t+\Delta t} = \mu_t + \kappa(\theta - \mu_t)\Delta t + \sigma\sqrt{\mu_t}(W_{t+\Delta t} - W_t). \tag{31}$$

However process (31) does not have a *strictly* zero probability of being negative as the Gaussian increment is not bounded from below and it is important to notice that the term $\sqrt{\mu_t}$ is not defined for a given $\mu_t < 0$. Hence in our numerical filtering procedures (application of algorithm described in section C), we use instead the following discretization:

$$\mu_{t+\Delta t} = \mu_t + \kappa(\theta - \mu_t)\Delta t + \sigma\sqrt{|\mu_t|}(W_{t+\Delta t} - W_t). \tag{32}$$

Berkaoui et al. (2008) show that for the process defined by (32), for all t given $\mu_{\eta(t)} = x$, where x is a positive value and $\eta(t)$ is the closest previous step of discretization, the following probability inequality applies

$$\mathbb{P}^x_{cir} = \mathbb{P}(\mu_t \leq 0 \mid \mu_{\eta(t)} = x) \leq \frac{1}{2}\exp\left(-\frac{x}{8\sigma^2\Delta t}\right); \tag{33}$$

thus, x being sufficiently above zero and/or taking time steps Δt sufficiently small, greatly lowers the chance of μ_t of becoming negative. Notice that, for the AR system, the value of the conditional probability $\mathbb{P}(\mu_t \leq 0 \mid \mu_{t-1} = x)$ has the following expression if we assume that μ has the AR(1) dynamics given by equation (2) in the text:

$$\mathbb{P}^x_{ar} = \mathbb{P}\left(w_t \leq -(1-\beta)E_r - \beta x\right) = \Phi\left(\frac{-(1-\beta)E_r - \beta x}{\sigma_{w,ar}}\right), \tag{34}$$

where Φ stands for the cumulative distribution function of a standard Gaussian random variable. For instance, setting $x = 0.5\%$ and considering parameter values used in section 2.4, (33) and (34) lead to $\mathbb{P}^{0.005}_{cir} \leq 5 \times 10^{-5}$ and $\mathbb{P}^{0.005}_{ar} = 0.22$. This result is confirmed visually by expected returns presented in Figure 2.

Throughout our economical analysis we assume that actual realizations of the process μ, given by equation (31), are non-negative. This is a reasonable assumption given the fact that μ is an unobservable process and that the non-negativity assumption is one of the two motivations for using a discretized CIR process instead of the original AR(1) process. On the other hand, if the aim would be to simulate the CIR-type process μ_t (instead of estimating it), a more adapted equation for this purpose would be for instance the symmetrized Euler scheme of (30), studied by Bossy, Diop, et al. (2007):

$$\hat{\mu}_{t+\Delta t} = \left|\hat{\mu}_t + \kappa(\theta - \hat{\mu}_t)\Delta t + \sigma\sqrt{\hat{\mu}_t}(W_{t+\Delta t} - W_t)\right|, \tag{35}$$

which ensures the positivity of the discretized process[21]. However discretization (35) does not allow to explore all the economic implications of the model, as we are unable to

[21] Alfonsi (2005) discusses several discretization schemes for the simulation of the CIR processes.

derive closed-form solutions for the autocovariance of returns (equation 10 in the text) and the variance of μ, for instance. Thus, in order to develop the economic implications of the expected return positivity condition on the predictive system we work with the CIR discretization (31) and assume that any actual path of μ remains positive in our analysis.

B Autocovariance of returns

In this section we derive the autocovariance of returns using equation (1) in the text. We assume first the following (more general) dynamics for μ:

$$\mu_{t+1} = (1-\beta)E_r + \beta\mu_t + g(\mu_t)w_{t+1}, \tag{36}$$

where g is a general function including the case $g(.) = \sqrt{(.)}$ or $g(.) = \sqrt{|.|}$.
Notice that the autocovariance of returns at lag $k \geq 0$ is:

$$\begin{aligned}
\mathrm{Cov}(r_{t+k}, r_t) &= \mathrm{Cov}(\mu_{t+k-1} + u_{t+k},\ \mu_{t-1} + u_t) \\
&= \mathrm{Cov}(\mu_{t+k-1}, \mu_{t-1}) + \mathrm{Cov}(\mu_{t+k-1}, u_t).
\end{aligned} \tag{37}$$

The first term on the right hand side of equation (37) is the autocovariance of expected returns at lag k. Using equation (8) (with $\sqrt{(.)}$ expressed as $g(.)$) the autocovariance of μ is equal to:

$$\begin{aligned}
\mathrm{Cov}(\mu_{t+k}, \mu_t) &= \mathrm{Cov}\left(\sum_{j=0}^{\infty} \beta^j w_{t+k-j} g(\mu_{t+k-1-j}),\ \sum_{i=0}^{\infty} \beta^i w_{t-i} g(\mu_{t-1-i})\right) \\
&= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} \beta^{i+j} \mathrm{Cov}\left(w_{t+k-j} g(\mu_{t+k-1-j}),\ w_{t-i} g(\mu_{t-1-i})\right) \\
&= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} \beta^{i+j} \mathbb{E}\left[w_{t+k-j} g(\mu_{t+k-1-j})\ w_{t-i} g(\mu_{t-1-i})\right].
\end{aligned}$$

To see this notice that the product of expectations is zero in the covariance due to the lag between the noise terms and the terms containing $g(\mu)$ (which are independent). Moreover, the only nonzero terms in the sums are when $j = i + k$. Thus we obtain:

$$\mathrm{Cov}(\mu_{t+k}, \mu_t) = \sum_{i=0}^{\infty} \beta^{2i+k} \sigma_w^2 \mathbb{E}\left[g(\mu_{t-1-i})^2\right] = \beta^k \sigma_w^2 \sum_{i=0}^{\infty} \beta^{2i} \mathbb{E}\left[g(\mu_{t-1-i})^2\right]. \tag{38}$$

Equation (38) easily simplifies to $\beta^k\left(\frac{\sigma_w^2 E_r}{1-\beta^2}\right)$ and allows us to compute the variance of μ when $g(.) = \sqrt{(.)}$ as in section 2.3 of the text. Note that this result is valid if the trajectory of μ stays positive at all time such that the term $\sqrt{\mu_t}$ is well defined for all t. However, as explained above, using a discretization of a CIR process (with a Gaussian noise) introduces a bias that could lead to eventual negative values (with low probability, cf. Appendix A) forcing us to consider the case $g(.) = \sqrt{|.|}$. In order to check that the use of this function $g(.)$ (defined for all real numbers) in our empirical analysis does not alter the analytical properties of the model, we derive here the autocovariance in this particular case which has to be linked with equations (10) and (11) of the text. The derivation of the second term of the right hand side of (37) yields to:

$$\begin{aligned}
\mathrm{Cov}(\mu_{t+k-1}, u_t) &= \mathrm{Cov}\left(\sum_{i=0}^{\infty} \beta^i w_{t+k-1-i}\sqrt{|\mu_{t+k-2-i}|},\ u_t\right) \\
&= \sum_{i=0}^{\infty} \beta^i \mathrm{Cov}(w_{t+k-1-i}\sqrt{|\mu_{t+k-2-i}|},\ u_t) \\
&= \sum_{i=0}^{\infty} \beta^i \mathbb{E}[w_{t+k-1-i}\sqrt{|\mu_{t+k-2-i}|}\ u_t],
\end{aligned}$$

where the only nonzero term is obtained when $i = k-1$. Thereby,

$$\mathrm{Cov}(\mu_{t+k-1}, u_t) = \beta^{k-1}\mathbb{E}[w_t\sqrt{|\mu_{t-1}|}\ u_t] = \beta^{k-1}\sigma_{uw}\mathbb{E}[\sqrt{|\mu_{t-1}|}]. \tag{39}$$

The autocovariance of μ is now:

$$\begin{aligned}
\mathrm{Cov}(\mu_{t+k}, \mu_t) &= \sum_{i=0}^{\infty} \beta^{2i+k}\sigma_w^2 \mathbb{E}|\mu_{t-1-i}| = \beta^k\sigma_w^2 \sum_{i=0}^{\infty} \beta^{2i}\left\{\mathbb{E}(\mu_{t-1-i}) + \mathbb{E}\left(|\mu_{t-1-i}| - \ \mu_{t-1-i}\right)\right\} \\
&= \beta^k\left(\frac{\sigma_w^2 E_r}{1-\beta^2}\right) + \beta^k\epsilon,
\end{aligned} \tag{40}$$

where $\epsilon = \sigma_w^2\sum_{i=0}^{\infty}\beta^{2i}\mathbb{E}\left(|\mu_{t-1-i}| - \ \mu_{t-1-i}\right)$ represents the negligeable imperfection error (cf. Proposition 1 below). This leads to:

$$\mathrm{Cov}(r_{t+k}, r_t) = \beta^{k-1}\left\{\beta\left(\frac{\sigma_w^2 E_r}{1-\beta^2}\right) + \sigma_{uw}\mathbb{E}(\sqrt{|\mu_{t-1}|})\right\} + \beta^k\epsilon. \tag{41}$$

Proposition 1 *Let the expected return process follow the equation,*

$$\mu_{t+1} = (1-\beta)E_r + \beta\mu_t + \sqrt{|\mu_t|}w_{t+1},$$

where $w_t \sim \mathcal{N}(0, \sigma_w)$, $E_r > 0$ and $\beta \in (0,1)$. The following result holds

$$\epsilon = \sigma_w^2 \sum_{i=0}^{\infty} \beta^{2i} \mathbb{E}\left(|\mu_{t-1-i}| - \ \mu_{t-1-i}\right) \to 0$$

as $\Delta t \to 0$, with $n \to \infty$ and $n\Delta t \to \infty$, where the time index $t+k$ (in the time series notation) corresponds to the time $(n+k)\Delta t$, for $k \in \{0, 1, 2...\}$. Moreover, the term β^k factor of ϵ in (41) further increases the convergence to zero.

Proof:

$$\begin{aligned}\epsilon &= \sigma_w^2 \sum_{i=0}^{\infty} \beta^{2i} \mathbb{E}\left(|\mu_{(n-1-i)\Delta t}| - \ \mu_{(n-1-i)\Delta t}\right) = -2\sigma_w^2 \sum_{i=0}^{\infty} \beta^{2i} \mathbb{E}\left(\mu_{(n-1-i)\Delta t}\right)^- \\ &= -2\beta^{2n}\sigma^2 \sum_{i=-\infty}^{n-1} \Delta t \beta^{-2(i+1)} \mathbb{E}\left(\mu_{i\Delta t}\right)^- ,\end{aligned}$$

where we use the fact that $\sigma_w^2 = \sigma^2 \Delta t$. This is of the form:

$$\kappa \mathbb{E}[\langle \nu^{n-1}, f \rangle],$$

where $f(x) = (x)^-$ and

$$\nu^m(\omega, dx) = \frac{1}{H^m} \sum_{i=-\infty}^{m} \beta^{-2(i+1)} \mathbb{I}_{\{\mu_{i\Delta t}(\omega) \in dx\}}$$

with

$$H^m = \sum_{i=-\infty}^{m} \beta^{-2(i+1)} = \beta^{-2(m+1)} \sum_{i=0}^{\infty} \beta^{2i} = \frac{\beta^{-2(m+1)}}{1-\beta^2} = \frac{\beta^{-2(m+1)}}{2\kappa\Delta t - \kappa^2 \Delta t^2} \leq \frac{\beta^{-2(m+1)}}{\kappa \Delta t},$$

where we have used the equality $\beta = (1 - \kappa\Delta t)$. Assume $\Delta t = \alpha(n) \to 0$, with $n \to +\infty$ and $n\Delta t \to +\infty$. Combining the weak convergence Theorem 2 of Pages, Panloup, et al. (2009) with the estimation of strong convergence of the symetrized Euler scheme using μ for X in Berkaoui et al. (2008), we get that for any continuous bounded function f

$$\frac{1}{H^n} \sum_{i=-\infty}^{n} \beta^{-2(i+1)} f(\mu_{i\Delta t}) \xrightarrow{n \to +\infty} \int_{\mathbb{R}+} f(x) \nu_0(dx) = 0, \ a.s.,$$

where ν_0 is the unique invariant measure of the CIR process (Gamma-type law). Due to the fact that the support of the function $f(x) = (x)^-$ is $\mathbb{R}^-$, and the support of a Gamma law is $(0, \infty)$, the integral and thus the limit of ϵ is equal to zero. $\square$

On the other hand, assuming the positivity of the previous value of μ and removing the absolute value within the square root, the derivation of the autocovariance of returns is simpler and yields to:

$$\mathrm{Cov}(r_{t+k}, r_t) = \beta^{k-1} \left\{ \beta \left(\frac{\sigma_w^2 E_r}{1-\beta^2} \right) + \sigma_{uw} \mathbb{E}(\sqrt{\mu_{t-1}}) \right\}. \tag{42}$$

According to Proposition 1, the autocovariance of the process with $g(.) = \sqrt{|.|}$ boils down to equation (42) for $\mu_{t-1} \geq 0$, which is equation (10) in the text.

C The Kalman filter

As mentioned earlier in the text, in this paper our analysis is performed without predictors, given that we focus on the implications arising from a modified interaction between past returns and expected returns. In this section we describe the procedure to estimate the unobservable process μ based on observations of both r and x, so that further works could be eventually conducted with predictors. The configuration investigated in the text corresponds simply to not consider the terms related to x in this section. In the next paragraph we present the algorithm for a state process $\{\mu_t\}$ with dynamics described by equation (36) above. To do this, we use an extended version of the Kalman filter (see Anderson and Moore, 2012, Chap. 8), to which we add the predictors $\{x_t\}$ in order to present the estimation procedure of the full system including eventual predictors. The Kalman filter theory relies on the assumption that, conditional to the information available at time $t-1$, denoted here D_{t-1}, i.e., $D_{t-1} = (r_1, x_1, r_2, x_2, ..., r_{t-1}, x_{t-1})$, the state variable μ_t has a Gaussian distribution. This assumption must also hold conditioned on D_t. In the configuration described above, using the extended Kalman filter consists in linearizing the function g around our last estimation of μ, i.e., replacing the term $g(\mu_t)$ by $g\left(\mathbb{E}(\mu_t|D_t)\right)$ in our procedure.

C.1 The algorithm

Following Pastor and Stambaugh (2009), we use the following notations:

$$z_t = \begin{bmatrix} r_t \\ x_t \end{bmatrix}, \quad a_t = \mathbb{E}(\mu_t|D_{t-1}), \quad b_t = \mathbb{E}(\mu_t|D_t), \quad f_t = \mathbb{E}(z_t|D_{t-1}), \quad P_t = \mathrm{Var}(\mu_t|D_{t-1}),$$

$$Q_t = \mathrm{Var}(\mu_t|D_t), \quad R_t = \mathrm{Var}(z_t|\mu_t, D_{t-1}), \quad S_t = \mathrm{Var}(z_t|D_{t-1}), \quad G_t = \mathrm{Cov}(z_t, \mu_t|D_{t-1}).$$

Initialization We assume conditioning on the (unknown) parameters even if not explicitly specified and that D_0 denotes the null information.
Assuming that $\mu_1 \sim \mathcal{N}(E_r, V_\mu)$ and $r_1 \sim \mathcal{N}(E_r, V_r)$, given $V_x, V_{rx}, V_{r\mu}, V_{x\mu}$, we have first

$$a_1 = E_r, \quad P_1 = V_\mu, \quad f_1 = [E_r \ E_x]', \quad S_1 = \begin{bmatrix} V_r & V_{rx} \\ V_{rx} & V_x \end{bmatrix}, \quad G_1 = [V_{r\mu} \ V_{x\mu}]',$$

$$\begin{aligned} R_1 &= S_1 - G_1 P_1^{-1} G_1', \\ Q_1 &= P_1 (P_1 + G_1' R_1^{-1} G_1)^{-1} P_1, \\ b_1 &= a_1 + P_1 (P_1 + G_1' R_1^{-1} G_1)^{-1} G_1' R_1^{-1} (z_1 - f_1). \end{aligned}$$

Iteration We use the extended Kalman filter algorithm to derive, for $t = 2, ..., T$,

$$a_t = (1-\beta)E_r + \beta \mathbb{E}(\mu_{t-1}|D_{t-1}) + \mathbb{E}(g(b_{t-1})w_t|D_{t-1}) = (1-\beta)E_r + \beta b_{t-1}. \tag{43}$$

$$\begin{aligned} P_t &= \mathrm{Var}((1-\beta)E_r + \beta\mu_{t-1} + g(b_{t-1})w_t|D_{t-1}) \\ &= \beta^2 \mathrm{Var}(\mu_{t-1}|D_{t-1}) + \mathrm{Var}(g(b_{t-1})w_t|D_{t-1}) + 2\beta \mathrm{Cov}(\mu_{t-1}, g(b_{t-1})w_t|D_{t-1}) \\ &= \beta^2 Q_{t-1} + g(b_{t-1})^2 \sigma_w^2. \end{aligned} \tag{44}$$

We have:

$$S_t = \begin{bmatrix} \mathrm{Var}(r_t|D_{t-1}) & \mathrm{Cov}(x_t, r_t|D_{t-1}) \\ \mathrm{Cov}(r_t, x_t|D_{t-1}) & \mathrm{Var}(x_t|D_{t-1}) \end{bmatrix} = \begin{bmatrix} Q_{t-1} + \sigma_u^2 & \sigma_{uv} \\ \sigma_{vu} & \Sigma_{vv} \end{bmatrix} \tag{45}$$

$$G_t = \begin{bmatrix} G_t^1 \\ G_t^2 \end{bmatrix},$$

with

$$\begin{aligned} G_t^1 &= \mathrm{Cov}(\mu_{t-1} + u_t, (1-\beta)E_r + \beta\mu_{t-1} + g(b_{t-1})w_t|D_{t-1}) \\ &= \beta Q_{t-1} + \mathrm{Cov}(\mu_{t-1}, g(b_{t-1})w_t|D_{t-1}) \\ &+ \beta \mathrm{Cov}(u_t, \mu_{t-1}|D_{t-1}) + \mathrm{Cov}(u_t, g(b_{t-1})w_t|D_{t-1}) \\ &= \beta Q_{t-1} + g(b_{t-1})\sigma_{uw}, \end{aligned}$$

and

$$
\begin{aligned}
G_t^2 &= \mathrm{Cov}((I-A)E_x + Ax_{t-1} + v_t, (1-\beta)E_r + \beta\mu_{t-1} + g(b_{t-1})w_t | D_{t-1}) \\
&= \mathrm{Cov}(v_t, g(b_{t-1})w_t | D_{t-1}) \\
&= g(b_{t-1})\sigma_{vw}.
\end{aligned}
$$

Finally,

$$
G_t = \begin{bmatrix} \beta Q_{t-1} + g(b_{t-1})\sigma_{uw} \\ g(b_{t-1})\sigma_{vw} \end{bmatrix}. \tag{46}
$$

The last terms are functions of those previously computed:

$$
R_t = S_t - G_t P_t^{-1} G_t', \tag{47}
$$

$$
Q_t = P_t(P_t + G_t' R_t^{-1} G_t)^{-1} P_t, \tag{48}
$$

$$
f_t = \begin{bmatrix} \mathbb{E}(\mu_{t-1}|D_{t-1}) \\ (I-A)E_x + Ax_{t-1} \end{bmatrix} = \begin{bmatrix} b_{t-1} \\ (I-A)E_x + Ax_{t-1} \end{bmatrix}. \tag{49}
$$

The filtering term b_t is given by

$$
b_t = a_t + P_t(P_t + G_t' R_t^{-1} G_t)^{-1} G_t' R_t^{-1}(z_t - f_t) = a_t + G_t' S_t^{-1}(z_t - f_t). \tag{50}
$$

As in Pastor and Stambaugh (2009), denote

$$
\begin{aligned}
[m_t\ n_t'] &= P_t(P_t + G_t' R_t^{-1} G_t)^{-1} G_t' R_t^{-1} = G_t' S_t^{-1} \\
&= \mathrm{Cov}(z_t', \mu_t | D_{t-1})[\mathrm{Var}(z_t|D_{t-1})]^{-1} \\
&= [\beta Q_{t-1} + g(b_{t-1})\sigma_{uw} \ \ g(b_{t-1})\sigma_{vw}] \begin{bmatrix} Q_{t-1} + \sigma_u^2 & \sigma_{uv} \\ \sigma_{vu} & \Sigma_{vv} \end{bmatrix}^{-1}.
\end{aligned} \tag{51}
$$

Notice that the terms defining m_t (and n_t) depend on $g(b_{t-1})$, which suggests a higher time dependence with respect to the terms in Pástor and Stambaugh (2009) setting. This implies that the level and change effect might have a more variable relative importance over time in our setting relative to the AR system.

From equation (50), we derive

$$
b_t = a_t + [m_t\ n_t'](z_t - f_t) \tag{52}
$$

$$
= (1-\beta)E_r + \beta b_{t-1} + [m_t\ n_t'] \begin{bmatrix} r_t - b_{t-1} \\ x_t - (I-A)E_x - Ax_{t-1} \end{bmatrix} \tag{53}
$$

$$
= (1-\beta)E_r + (\beta - m_t)b_{t-1} + m_t r_t + n_t' v_t. \tag{54}
$$

By repeated substitutions of the lagged values of $(b_t - E_r)$ in equation (54) we obtain:

$$b_t = E_r + \sum_{s=1}^{t} \lambda_s (r_s - b_{s-1}) + \sum_{s=1}^{t} \phi_s' v_s, \tag{55}$$

where $\lambda_s = m_s \beta^{t-s}$ and $\phi_s = n_s \beta^{t-s}$ and $(r_s - b_{s-1}) = r_t - \mathbb{E}(r_t | D_{t-1})$ is the forecast error. Equation (55) has the same structure than the equivalent expression in the original predictive system of Pástor and Stambaugh (2009), but the coefficients m_s and n_s are modified, leading to different predictions. Equation (55) can be rewritten as a function of past returns instead of past forecast errors as follows

$$b_t = E_r + \sum_{s=1}^{t} \omega_s (r_s - E_r) + \sum_{s=1}^{t} \delta_s' v_s, \tag{56}$$

where,

$$\omega_s = \begin{cases} (\beta - m_t)(\beta - m_{t-1}) \dots (\beta - m_{s+1}) m_s & , \text{ for } s < t \\ m_s & , \text{ for } s = t. \end{cases}$$

and

$$\delta_s = \begin{cases} (\beta - m_t)(\beta - m_{t-1}) \dots (\beta - m_{s+1}) n_s & , \text{ for } s < t \\ n_s & , \text{ for } s = t. \end{cases}$$

If E_r is replaced by the sample mean in equation (56), it can be shown that the estimate of b_t is

$$\hat{b}_t = \sum_{s=1}^{t} \kappa_s r_s + \sum_{s=1}^{t} \delta_s' v_s, \tag{57}$$

where

$$\kappa_s = \frac{1}{t} \left(1 - \sum_{l=1}^{t} \omega_l \right) + \omega_s, \tag{58}$$

and $\sum_{s=1}^{t} \kappa_s = 1$. This expression has the same form than in the original predictive system of Pástor and Stambaugh (2009), but the ω_s are functions of m_s, which has a different expression in our setting that depends on the level of μ_s. To see this, develop equation (54), add and subtract $m_t E_r$, rearrange terms and do backward substitution of $(b_t - E_r)$.

C.2 Steady state

Important results can be obtained assuming the system reach a steady state on the long run. Note that the results of Pages et al. (2009) ensure the existence of a steady state in the CIR discretization case, i.e., $g(.) = \sqrt{|.|}$. The expressions of the different elements defined at the beginning of section C can be derived, at the equilibrium, removing the subscripts t and $t-1$ of equations (44) to (48). We obtain the following system to solve for Q, the steady-state value of Q_t:

$$
\begin{aligned}
&P = \beta^2 Q + g(b)^2\sigma_w^2, \quad S = \begin{bmatrix} Q+\sigma_u^2 & \sigma_{uv} \\ \sigma_{vu} & \Sigma_{vv} \end{bmatrix}, \\
&G = \begin{bmatrix} \beta Q + g(b)\sigma_{uw} \\ g(b)\sigma_{vw} \end{bmatrix}, \quad R = S - GP^{-1}G', \\
&Q = P(P + G'R^{-1}G)^{-1}P.
\end{aligned}
$$

After rearranging terms, this gives us the following quadratic equation for Q:

$$Q^2 + \xi_1 Q + \xi_2 = 0, \tag{59}$$

where

$$
\begin{aligned}
\xi_1 &= (1-\beta^2)(\sigma_u^2 - \sigma_{uv}\Sigma_{vv}^{-1}\sigma_{vu}) + 2g(b)\beta(\sigma_{uw} - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vu}) - g(b)^2(\sigma_w^2 - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vw}) \\
&= (1-\beta^2)\mathrm{Var}(u \mid v) + 2g(b)\beta\mathrm{Cov}(u, w \mid v) - g(b)^2\mathrm{Var}(w \mid v),
\end{aligned} \tag{60}
$$

and

$$
\begin{aligned}
\xi_2 &= g(b)^2\left((\sigma_{uw} - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vu})^2 - (\sigma_u^2 - \sigma_{uv}\Sigma_{vv}^{-1}\sigma_{vu})(\sigma_w^2 - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vw})\right) \\
&= g(b)^2\left(\mathrm{Cov}(u, w \mid v)^2 - \mathrm{Var}(u \mid v)\mathrm{Var}(w \mid v)\right).
\end{aligned} \tag{61}
$$

The solution is thus the positive root of (59):

$$Q = \frac{\sqrt{\xi_1^2 - 4\xi_2} - \xi_1}{2}. \tag{62}$$

Moreover, using the value of Q given by (62) and equation (51), we obtain the steady-state expressions of m and n:

$$m = (\beta Q + g(b)\mathrm{Cov}(u, w \mid v))(Q + \mathrm{Var}(u \mid v))^{-1}, \tag{63}$$

$$n' = (g(b)\sigma_{wv} - m\sigma_{uv})\Sigma_{vv}^{-1}. \tag{64}$$

D Bayesian procedure

This section describes the Bayesian analysis of the CIR predictive system. As in Appendix C, we provide a description of the procedure for the full system, i.e., with eventual predictors. As Pástor and Stambaugh (2009), we use an MCMC procedure to obtain the posterior distribution of μ and θ the set of parameters, based on D the data available to the investor. We alternate between drawing μ from the posterior distribution $p(\mu|\theta, D)$ and drawing the parameters θ from the posterior $p(\theta|\mu, D)$.

D.1 Drawing μ_t

Given a set of parameters, we draw the time series of $\{\mu_t\}$ using the forward filtering, backward sampling approach of Carter and Kohn (1994) and Frühwirth-Schnatter (1994). The first stage consists in applying the Kalman filter procedure described above in section C, with the current set of parameters. The sampling stage is the same as described in section B3.2. of the internet appendix of Pástor and Stambaugh (2009). However, given that one of the motivations to use a modified version of the original AR predictive system is the belief that $\mu_t > 0$, we choose to use a rejection sampling methodology, i.e., we impose to each draw of μ_t to be positive at all times. Our procedure is the following, at each time step μ_t is drawn using the distribution $\mu_t|\mu_{t+1}, D_t$ which is Gaussian (due to the use of the extended Kalman filter) and thus can lead to eventual negative values (though rare, we have to consider this eventuality due to the discretization imperfection). In the case where a negative value is drawn at a specific time step t^* for μ_{t^*}, we redraw μ_{t^*} using the same distribution until a positive values is obtained, i.e., we draw μ_t using an acceptance-rejection method[22].

D.2 Prior distributions

As mentioned in the text, the priors used in the case of the CIR system are very similar to those in the AR system used by Pástor and Stambaugh (2009). Thus, we refer to its internet appendix for a detailed description of the prior distributions. A brief description,

[22] In the case where after a maximum number of 500 trials μ_{t^*} is still negative we reject the set of parameters and the current draw of μ. We draw a new set of parameters to sample a whole new time series of μ_t. The percentage of parameters rejection is relatively small in all configurations we tested. For instance, using the benchmark prior on R^2, the average rejection rate are 2.64% for the more informative prior, and 3.3% for the noninformative prior, on the longest out-of-sample period.

and a discussion of the slight modifications of E_r prior due to the specification of the CIR system, are done in section 3.1 of the text.

D.3 Posterior distributions

Conditional on the current draw of $\{\mu_t\}$, the posterior distributions of the parameters are the same as describe in section B5.2. of the internet appendix of Pástor and Stambaugh (2009), except the terms affected by the new dynamics of μ described below, the others remain unchanged. In this section we denote as K the number of predictors (notice that in the results presented in the text $K = 0$ as no predictor is used) and T the last period at which returns are available for the estimation period of concern. We use $E_{x\mu 0}$ and $V_{x\mu 0}$ as notations for the prior mean and variance of the vector $[E_x\ E_r]'$. Let $\Sigma_{(vw)} = \begin{bmatrix} \Sigma_{vv} & \sigma_{vw} \\ \sigma_{wv} & \sigma_w^2 \end{bmatrix}$ and $q_t = \begin{bmatrix} x_t \\ \mu_t \end{bmatrix}$ for $t \in \{1, 2, ..., T\}$.

Posterior of $\mathbf{E_x}$ and $\mathbf{E_r}$ The posterior for $E_{x\mu} = [E_x\ E_r]'$ is still normal $E_{x\mu} \sim \mathcal{N}(\tilde{E}_{x\mu}, \tilde{V}_{x\mu})$ but with[23]

$$\tilde{V}_{x\mu} = \left(V_{x\mu 0}^{-1} + \sum_{t=1}^{T-1} L_2' \Sigma_{(vw)}^{g(\mu_t)} L_2 \right)^{-1} \quad \text{and} \quad \tilde{E}_{x\mu} = \tilde{V}_{x\mu} \left(V_{x\mu 0}^{-1} E_{x\mu 0} + L_2' \sum_{t=1}^{T-1} \Sigma_{(vw)}^{g(\mu_t)} \left(q_{t+1} - L_1 q_t \right) \right),$$

where $L_1 = \begin{bmatrix} A & 0 \\ 0 & \beta \end{bmatrix}$, $L_2 = \begin{bmatrix} I_K - A & 0 \\ 0 & 1-\beta \end{bmatrix}$ and $\Sigma_{(vw)}^{g(\mu_t)} = \Sigma_{(vw)} \circ \begin{bmatrix} \mathbf{1}_{K\times K} & g(\mu_t)\mathbf{1}_{K\times 1} \\ g(\mu_t)\mathbf{1}_{1\times K} & g(\mu_t)^2 \end{bmatrix}$.

Posterior of A and $\boldsymbol{\beta}$ The posterior of $\mathbf{b} = [\mathrm{vec}(A')\ \beta]'$ is still the Gaussian distribution described in section B.5.2.1. of internet appendix of Pástor and Stambaugh (2009). Using the same notation we have:

$$z = Z\mathbf{b} + \text{errors},$$

but the covariance matrix of the error terms is instead

$$\left(\Sigma_{(vw)} \otimes I_{T-1} \right) \circ \Lambda_{1,T-1},$$

[23]The operator $\circ$ denotes the Hadamard product, also known as the element-wise product. I_K is the identity matrix with dimension $K \times K$. For integers m and n, $\mathbf{1}_{m\times n}$ denotes a matrix with m rows and n columns whose all entries are 1.

where $\Lambda_{1,T-1}$ is the following $[(T-1)(K+1)] \times [(T-1)(K+1)]$ matrix:

$$\begin{bmatrix} \mathbf{1}_{K(T-1)\times K(T-1)} & \begin{matrix} \begin{bmatrix} g(\mu_1) & & \mathbf{1} \\ & \ddots & \\ \mathbf{1} & & g(\mu_{T-1}) \end{bmatrix} \\ \vdots \\ \begin{bmatrix} g(\mu_1) & & \mathbf{1} \\ & \ddots & \\ \mathbf{1} & & g(\mu_{T-1}) \end{bmatrix} \end{matrix} \\ \begin{bmatrix} g(\mu_1) & & \mathbf{1} \\ & \ddots & \\ \mathbf{1} & & g(\mu_{T-1}) \end{bmatrix} \quad \cdots \quad \begin{bmatrix} g(\mu_1) & & \mathbf{1} \\ & \ddots & \\ \mathbf{1} & & g(\mu_{T-1}) \end{bmatrix} & \begin{bmatrix} g(\mu_1)^2 & & \mathbf{1} \\ & \ddots & \\ \mathbf{1} & & g(\mu_{T-1})^2 \end{bmatrix} \end{bmatrix},$$

which is equal to the matrix of the AR system except for the terms involving μ (i.e., bottom and right block of the matrix above).

Posterior of $\boldsymbol{\Sigma}$ We follow the decomposition of Pástor and Stambaugh (2009) by changing variables from $\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{uw} \\ \sigma_{vu} & \Sigma_{vv} & \sigma_{vw} \\ \sigma_{wu} & \sigma_{wv} & \sigma_w^2 \end{bmatrix}$ to the set of (Σ_{11}, C, Ω), where $\Sigma_{11} = \begin{bmatrix} \sigma_u^2 & \sigma_{uw} \\ \sigma_{wu} & \sigma_w^2 \end{bmatrix}$ and C and Ω are the slope and the residual covariance matrix of the regression of v on (u, w). The procedure to draw (Σ_{11}, C, Ω) remains the same, except that given a draw of the time series of $\{\mu_t\}$, and conditional on (E_x, A, E_r, β), the sample of residuals of $\{w_t, t = 2, ..., T\}$ is no longer the time series $\{\mu_t - (1-\beta)E_r - \beta\mu_{t-1}, t = 2, ..., T\}$ but $\left\{\frac{\mu_t-(1-\beta)E_r-\beta\mu_{t-1}}{g(\mu_{t-1})}, t = 2, ..., T\right\}$. Thus using $g(.) = \sqrt{|.|}$ can lead to very large observations of w in magnitude when the current value of μ_{t-1} is close to zero. Thereby to avoid using biased estimators of variances and covariances involving w, we use a robust estimator to measure these terms. Specifically, let X denote the $(T-1) \times 2$ matrix of $[u_t, w_t]$ for $t = 2, ..., T$. Instead of using the classical estimator of variance for Σ_{11}: $\hat{\Sigma}_{11} = \frac{1}{T-1}(X'X)$, we use the robust and widely used *Minimum Covariance Determinant Estimator* (Fast MCD) of Rousseeuw and Driessen (1999) to compute $\hat{\Sigma}_{11}$. This method gives a weight vector ω of size $(T-1) \times 1$ with entries 0 or 1 for each observation based on a Mahalanobis distance criterium. Hence, in order to compute the parameters of the regression of v on (u, w) and derive the posterior distributions parameters of (Σ_{11}, C, Ω) described in section B5.2.2. of the internet appendix of Pástor and Stambaugh (2009), we use the corresponding reweighted vector $[\omega\ \omega] \circ X$ instead of X.

E Tables and Figures

R^2	E_r	σ_r	β	k-e ρ_{uw}	σ_μ	$\sigma_{w,ar}$	$\sigma_{w,cir}$	σ_u
0.09%	1.77%	8.38%	0.90	−0.06	0.25%	0.11%	0.82%	8.38%
5%	1.77%	8.38%	0.90	−0.47	1.87%	0.82%	6.14%	8.17%

Table 1: Point estimate parameters for different levels of the R^2 of the regression of r_{t+1} on μ_t. E_r and σ_r correspond to mean and standard deviation sample estimates for quarterly returns of the CRSP aggregate US market index from 1952 to 2012, and β is taken as in Pástor and Stambaugh (2009) section II.B. The following columns follow from the respective definitions in the text which are functions of the first four columns; k-e ρ_{uw} stands for knife-edge value of ρ_{uw}. Given a value for R^2 and assuming $\mathbb{E}(\sqrt{\mu_{t-1}}) \approx \sqrt{E_r}$, the knife-hedge value of ρ_{uw} is the same for both systems.

1975-2012	R^2_{OS}	β	R^2	ρ_{uw}
	AR system			
Best configuration	3.00%	0.9	4%	-0.95
Worst configuration	-5.81%	0.9	5%	0.95
	CIR system			
Best configuration	3.39%	0.8	2%	-0.95
Worst configuration	-5.62%	0.8	5%	0.95

Table 2: Out-of-sample results summary with point estimate parameters. Each line presents the highest (or lowest) R^2_{OS} obtained for each predictive system (AR or CIR), as well as the corresponding parameter set. Predictions are computed on quarterly returns on the value-weighted portfolio of all NYSE, Amex, and Nasdaq stocks in excess of the quarterly return on a 1-month T-bill obtained from CRSP. The sample begins in 1952 and the out-of-sample period is 1975-2012. We use the prevailing returns average for E_r.

1975-2012	Less predictability		Benchmark prior		More predictability	
Prior on ρ_{uw}	Noninf.	More inf.	Noninf.	More inf.	Noninf.	More inf.
CIR R^2_{OS} (%)	-0.30	1.21**	-0.73	1.07**	-1.26	0.71*
AR R^2_{OS} (%)	-1.03	0.28	-1.54	0.13	-2.23	-0.23
1975-2000	**Less predictability**		**Benchmark prior**		**More predictability**	
Prior on ρ_{uw}	Noninf.	More inf.	Noninf.	More inf.	Noninf.	More inf.
CIR R^2_{OS} (%)	-0.02	2.64***	-0.29	2.93***	-0.66	3.01***
AR R^2_{OS} (%)	-1.41	0.69*	-2.14	0.61*	-3.13	0.18
1980-2005	**Less predictability**		**Benchmark prior**		**More predictability**	
Prior on ρ_{uw}	Noninf.	More inf.	Noninf.	More inf.	Noninf.	More inf.
CIR R^2_{OS} (%)	0.41	1.48**	0.11	1.51**	-0.29	1.34**
AR R^2_{OS} (%)	-0.12	0.52	-0.34	0.62*	-0.72	0.66*
1985-2010	**Less predictability**		**Benchmark prior**		**More predictability**	
Prior on ρ_{uw}	Noninf.	More inf.	Noninf.	More inf.	Noninf.	More inf.
CIR R^2_{OS} (%)	0.04	0.80*	-0.41	0.67	-0.96	0.30
AR R^2_{OS} (%)	-0.20	0.04	-0.39	0.04	-0.73	0.02

Table 3: Out-of-sample results of the predictive systems using the Bayesian procedure. For each out-of-sample period, for both predictive systems, three priors on the R^2 of the regression of r_{t+1} on μ_t (less predictability, benchmark prior and more predictability) and two priors on ρ_{uw} (noninformative and more informative) are applied. Predictions are computed on quarterly returns on the value-weighted portfolio of all NYSE, Amex, and Nasdaq stocks in excess of the quarterly return on a 1-month T-bill obtained from CRSP. Our sample begins in 1952-Q1. The predictive systems parameters are re-estimated on the first available date of each year in the sample. Predictions are computed each new quarter using the data available at each point in time.

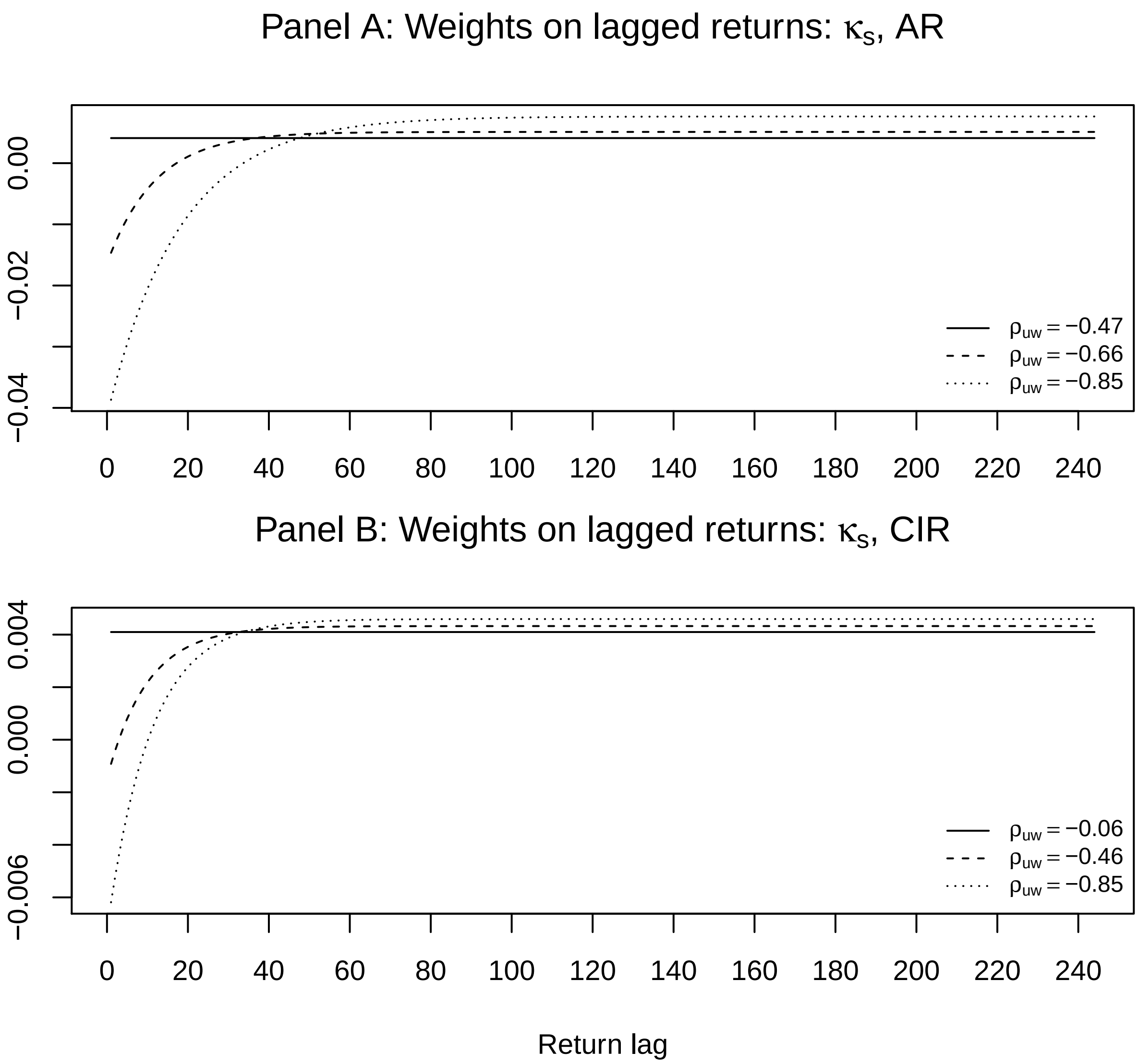

Figure 1: The effect of lagged returns on $E(r_{t+1}|D_t)$ when no predictors are used. This figures plots κ_s, the weights on lagged total returns in $E(r_{t+1}|D_t)$ when the unconditional mean return is estimated by the sample mean over the 244 quarters from 1952Q1 to 2012Q4. The different lines correspond to different values of ρ_{uw}, the correlation between expected and unexpected returns (the flat line corresponds to the knife-hedge value of ρ_{uw}, i.e., historical average as estimate of expected return). The autoregressive coefficient is set to $\beta = 0.9$. In Panel A, corresponding to Pástor and Stambaugh (2009)'s predictive system with AR(1) expected returns, the predictive R^2 corresponding to the fraction of variation in r_{t+1} that can be explained by μ_t is set to 0.05. In Panel B, the CIR-based method estimates are calculated using the same β and σ_w.

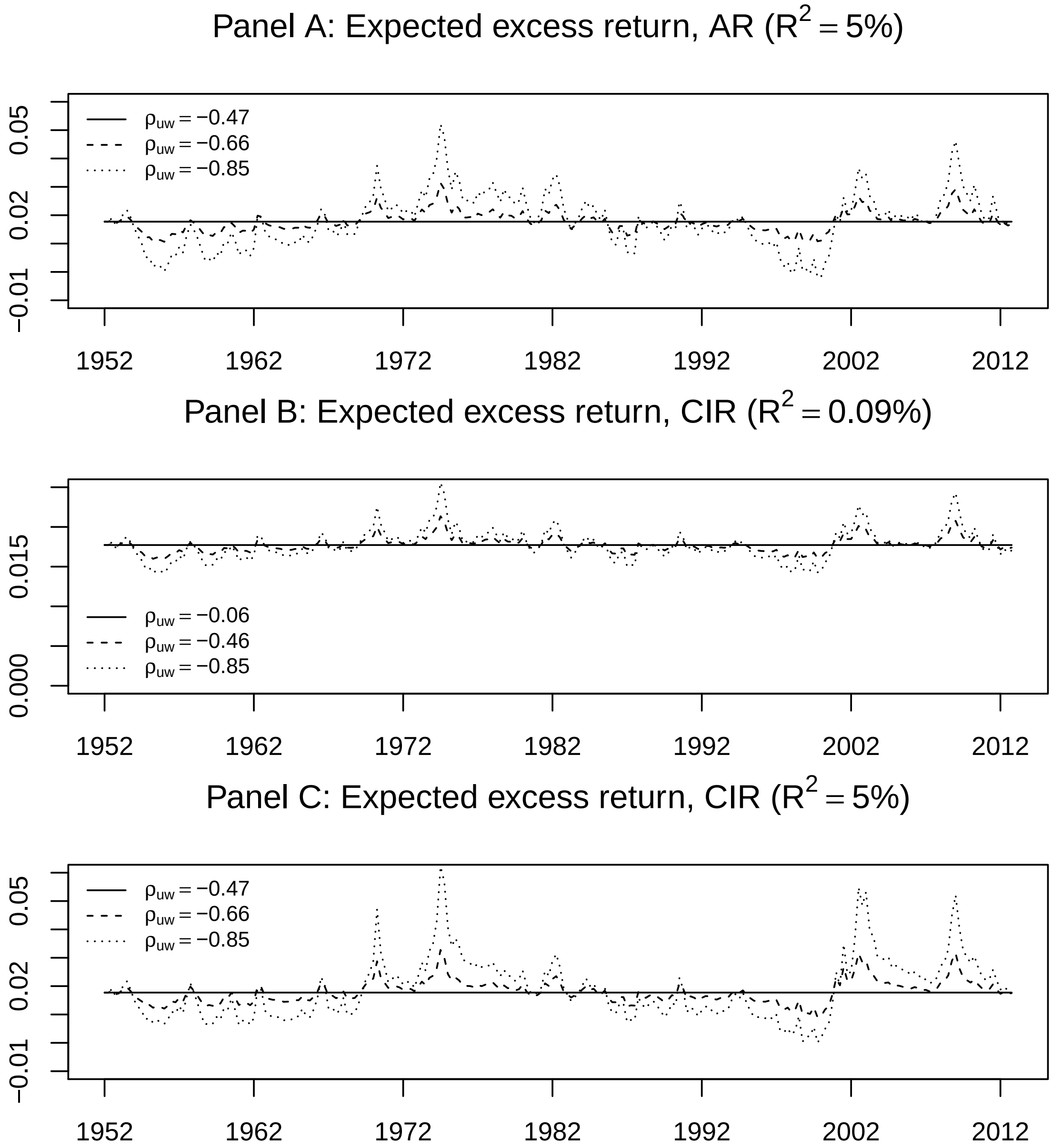

Figure 2: Panel A presents the equity premium $\mathbb{E}(r_{t+1}|D_t) = \mathbb{E}(\mu_t|D_t)$ from Pastor and Stambaugh's AR(1) predictive system, and Panel B and C the expected excess returns from our CIR-type predictive system. This figures displays the time series of quarterly US stock market premium from 1952Q1 to 2012Q4 estimated for different values of ρ_{uw} (the flat line corresponds to the knife-hedge value of ρ_{uw}, i.e., historical average as estimate of expected return). The autoregressive coefficient is set to $\beta = 0.9$ and the unconditional mean return E_r is estimated by the sample mean over the whole sample. In Panel A, the R^2 corresponding to the fraction of variation in r_{t+1} that can be explained by μ_t is set to 5%. In Panel B, the CIR-based method estimates are calculated using the same β and σ_w as in Panel A, leading to an R^2 equal to 0.09%. In Panel C, the R^2 is set to 5%.

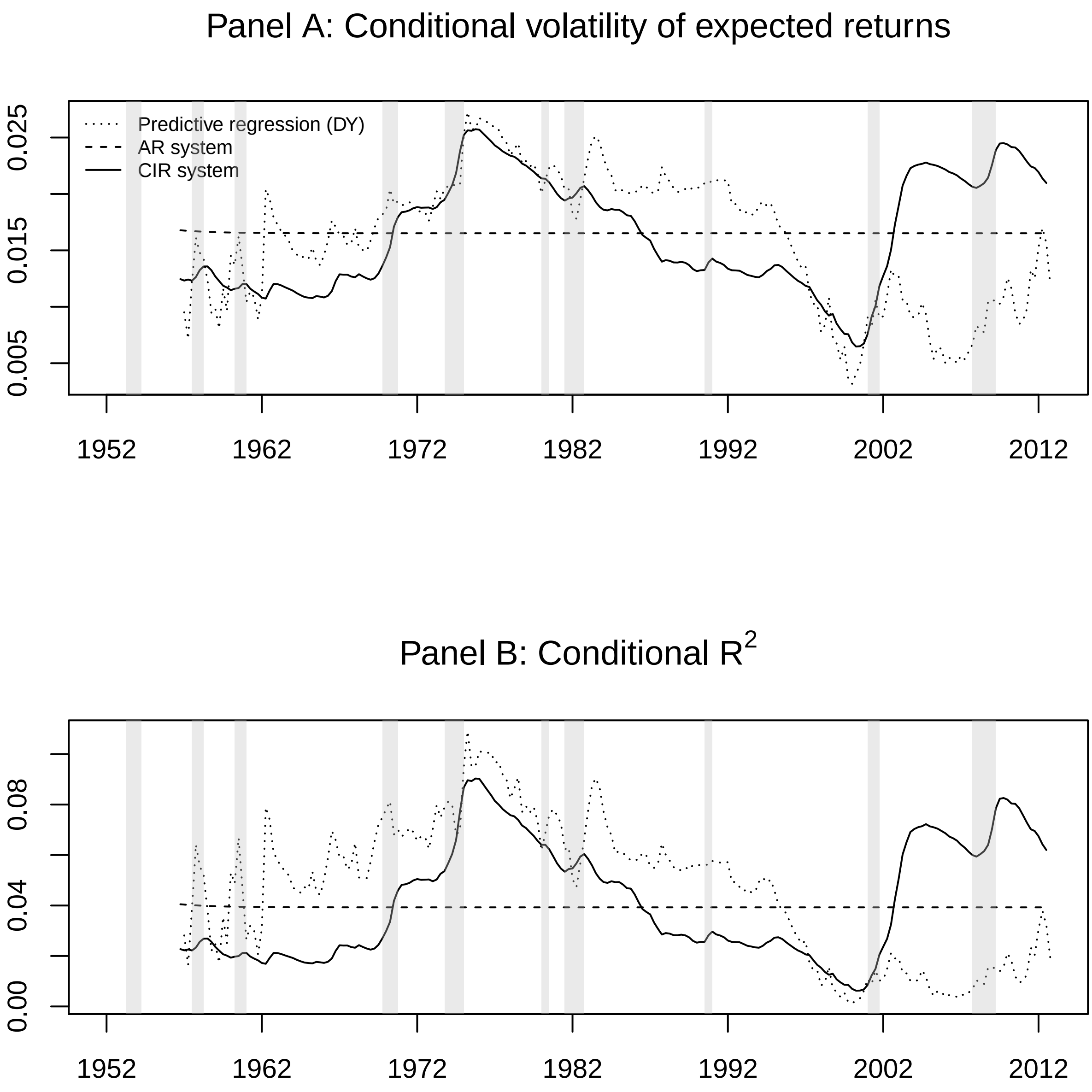

Figure 3: Panel A plots the square root of estimates of $\mathrm{Var}(\mu_t|D_t)$, where μ_t denotes the expected stock return from time t to time $t+1$ and D_t denotes the information set observed through time t. The conditional variance of μ is presented for the CIR system and the AR system without predictors and for the predictive regression using the dividend price-ratio as predictor. The sample considered is 1952Q1-2012Q4. The parameters used for the predictive systems corresponds to $R^2 = 5\%$, $\rho_{uw} = -0.85$ and $\beta = 0.9$. The unconditional mean E_r is estimated with the sample mean over the whole period. Panel B presents the ratio of the conditional variances of μ_t and r_{t+1} for both systems and the predictive regression. The grayed areas correspond to economic recessions as reported by NBER.

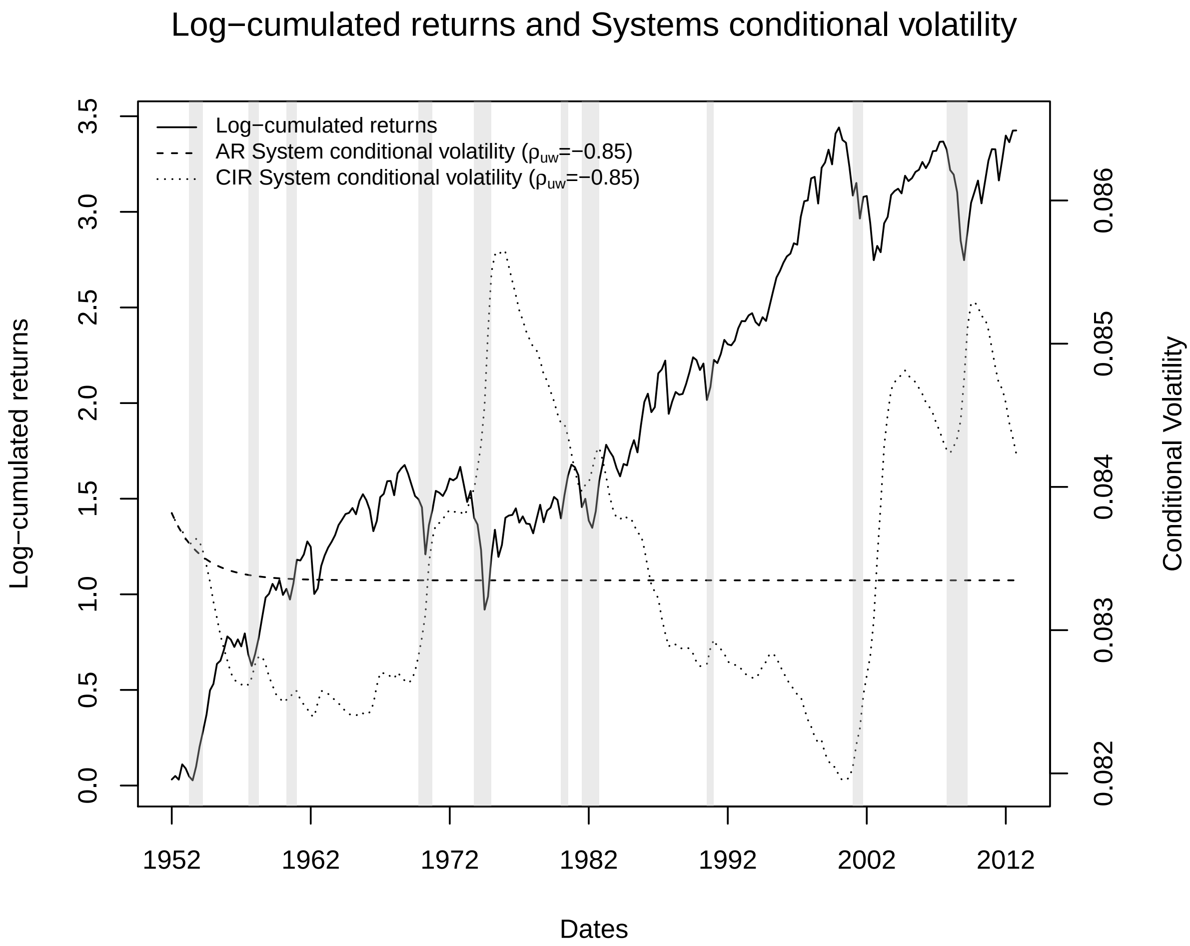

Figure 4: The log-cumulated returns and the conditional volatility of r_t when no predictors are used. This figure plots the cumulated returns of the market (log-value) and the square root of estimated values of $S_t = \mathrm{Var}(r_t|D_{t-1})$, where r_{t+1} denotes the stock return from time t to time $t+1$ and D_t denotes the history of returns observed through time t. The sample considered is 1952Q1-2012Q4. The conditional volatility is estimated using the CIR and AR predictive systems with $\rho_{uw} = -0.85$. The autoregressive coefficient is set to $\beta = 0.9$ and the unconditional mean return E_r is estimated by the sample mean over the whole sample. The predictive R^2 corresponding to the fraction of variation in r_{t+1} that can be explained by μ_t is set to 5% in both systems. The grayed areas correspond to economic recessions as reported by NBER.

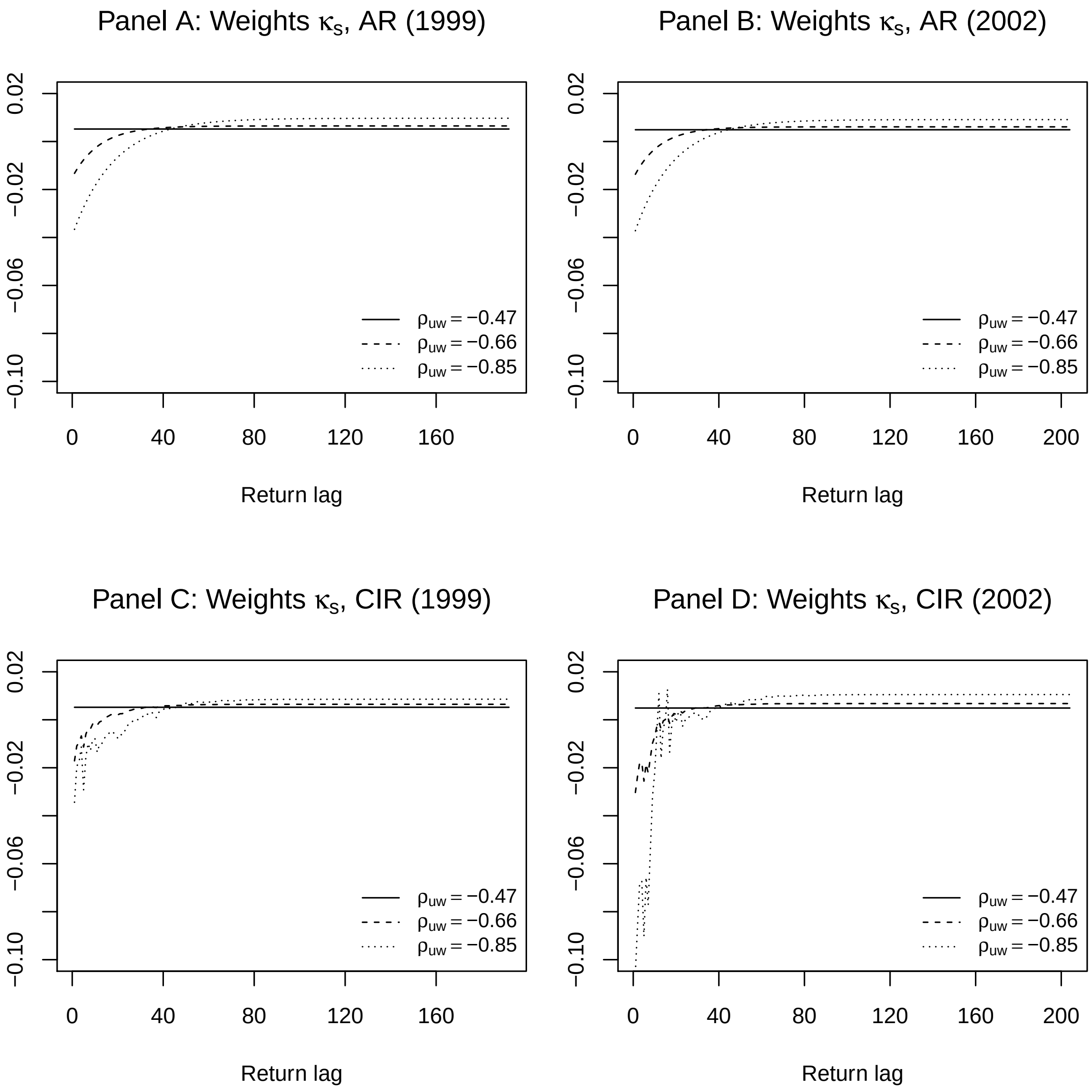

Figure 5: The effect of lagged returns on $E(r_{t+1}|D_t)$ when no predictors are used. This figures plots finite-sample values of κ_s, the weights on lagged total returns in $E(r_{t+1}|D_t)$. The samples considered are 1952Q1-1999Q4 for Panels A and C, and 1952Q1-2002Q4 for Panels B and D. The autoregressive coefficient is set to $\beta = 0.9$ and the unconditional mean return E_r is estimated by the sample mean over the quarters since 1952Q1. In both Panels, the predictive R^2 corresponding to the fraction of variation in r_{t+1} that can be explained by μ_t is set to 0.05.

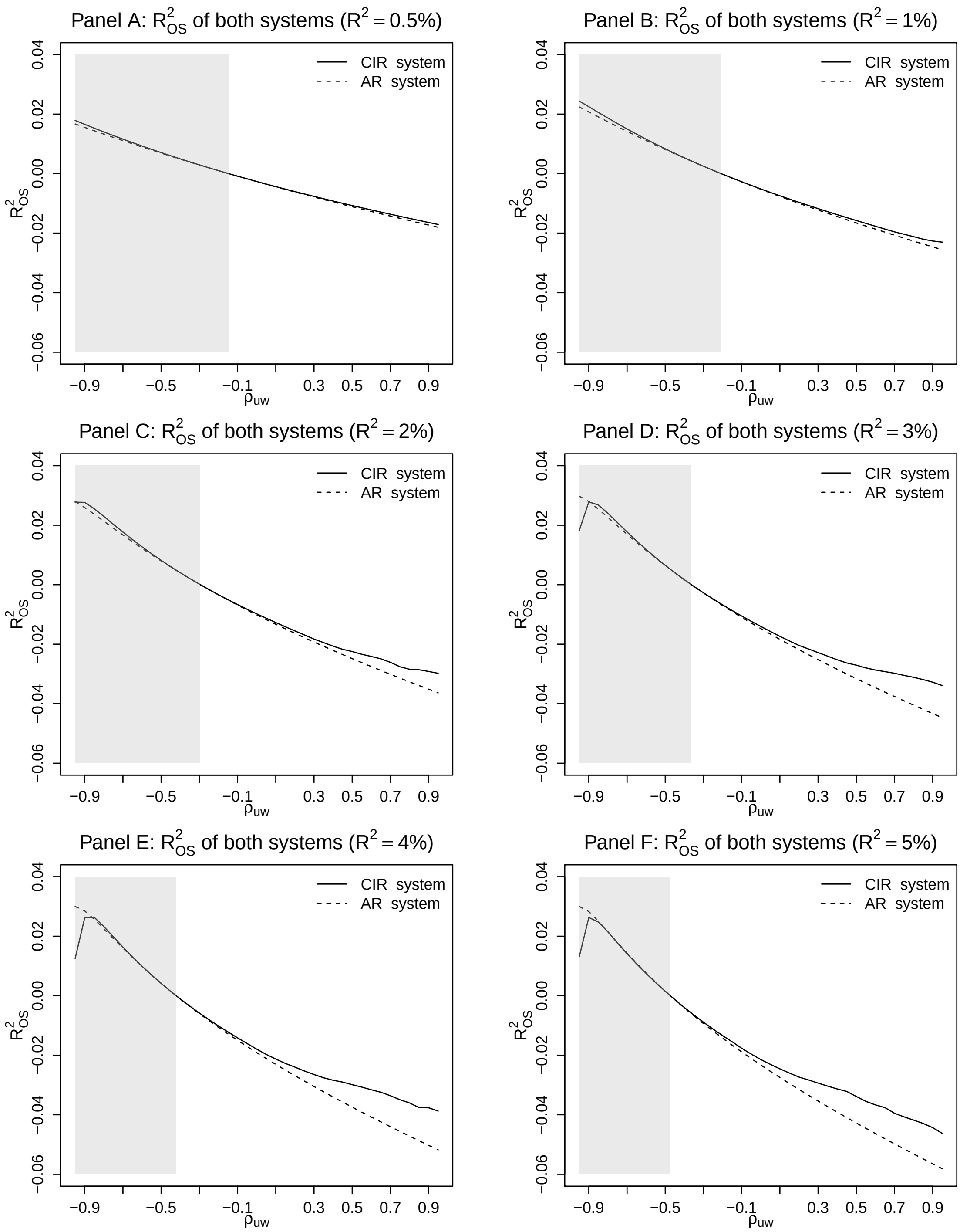

Figure 6: Out-of-sample results with point estimate parameters. Each Panel presents the R^2_{OS} as a function of ρ_{uw} when the R^2 of the regression of r_{t+1} on μ_t is set to a given value. Predictions are computed on quarterly returns on the value-weighted portfolio of all NYSE, Amex, and Nasdaq stocks in excess of the quarterly return on a 1-month T-bill obtained from CRSP. The sample begins in 1952 and the out-of-sample period is 1975-2012. We use a constant β of 0.9 and the prevailing returns average for E_r. The grayed areas correspond to values of ρ_{uw} implying a dominant change effect (countercyclical expected returns).

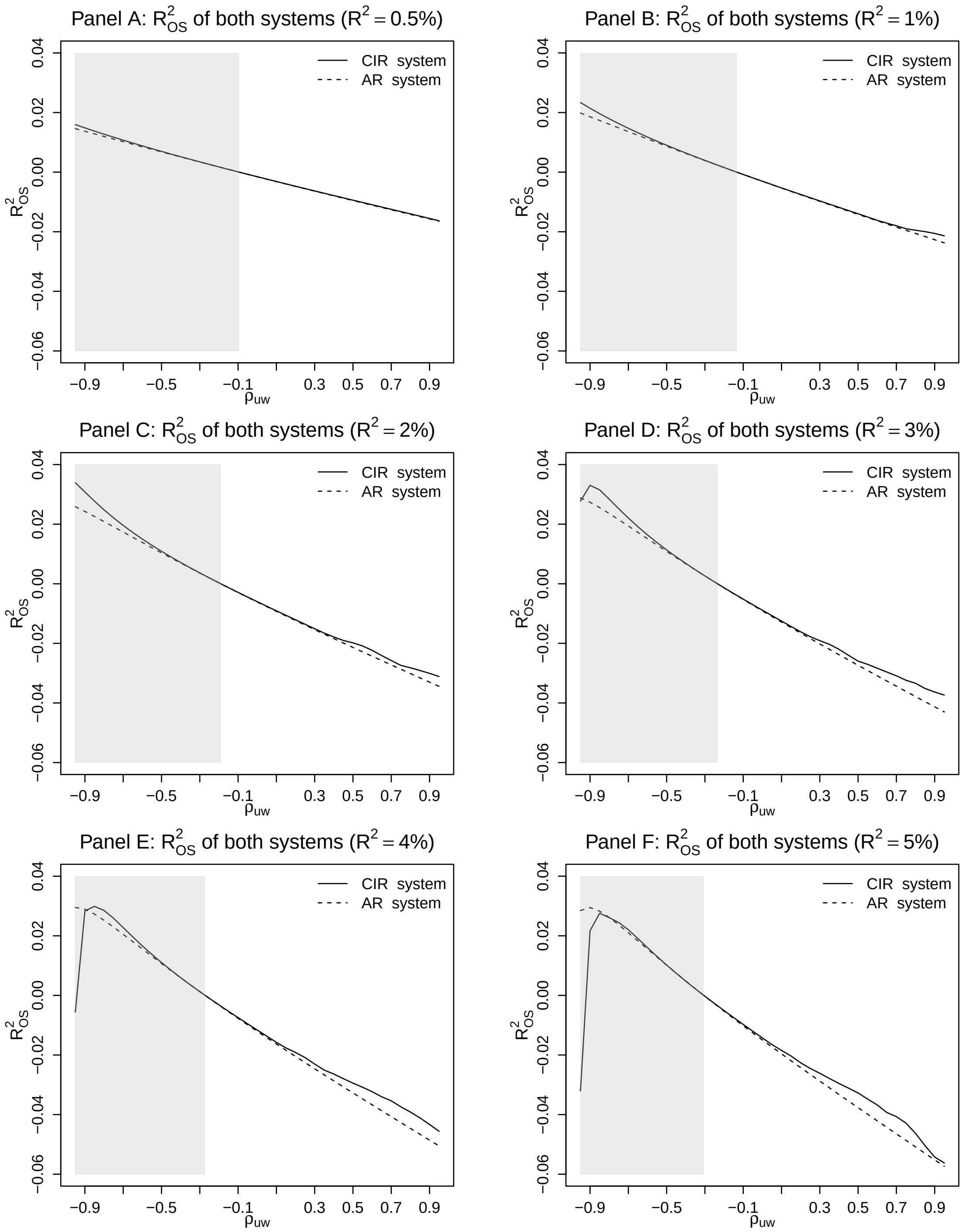

Figure 7: Out-of-sample results with point estimate parameters. Each Panel presents the R^2_{OS} as a function of ρ_{uw} when the R^2 of the regression of r_{t+1} on μ_t is set to a given value. Predictions are computed on quarterly returns on the value-weighted portfolio of all NYSE, Amex, and Nasdaq stocks in excess of the quarterly return on a 1-month T-bill obtained from CRSP. The sample begins in 1952 and the out-of-sample period is 1975-2012. We use a constant β of 0.8 and the prevailing returns average for E_r. The grayed areas correspond to values of ρ_{uw} implying a dominant change effect (countercyclical expected returns).

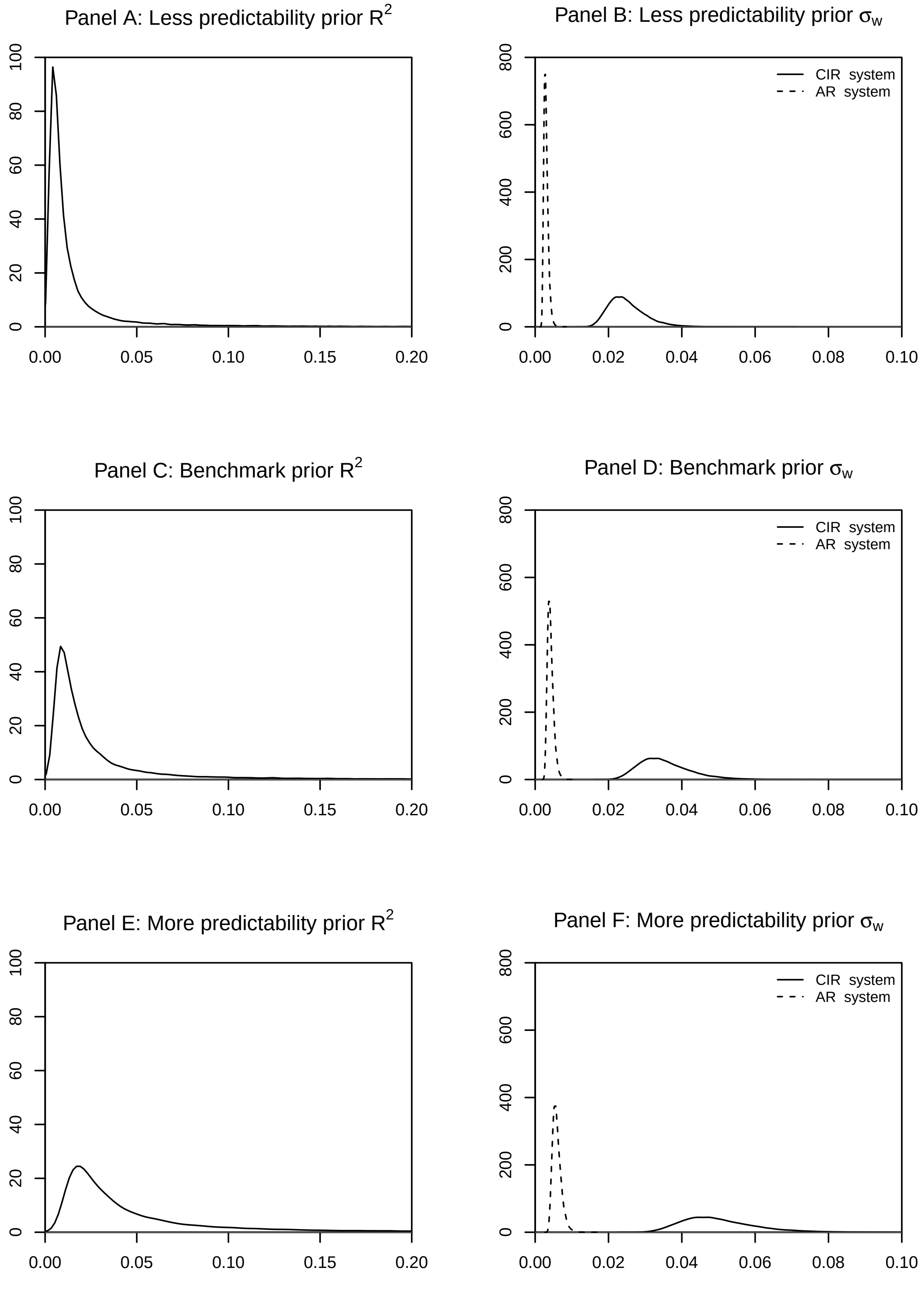

Figure 8: The prior distributions for R^2 and σ_w used in the Bayesian analysis. Panels A and B plot the prior on the R^2 from the regression of r_{t+1} on μ_t and the corresponding prior on σ_w for both AR system and CIR system, corresponding to less predictability prior. Panels C and D plot the distributions corresponding to the benchmark prior. Panels E and F plot the more predictability prior distributions.

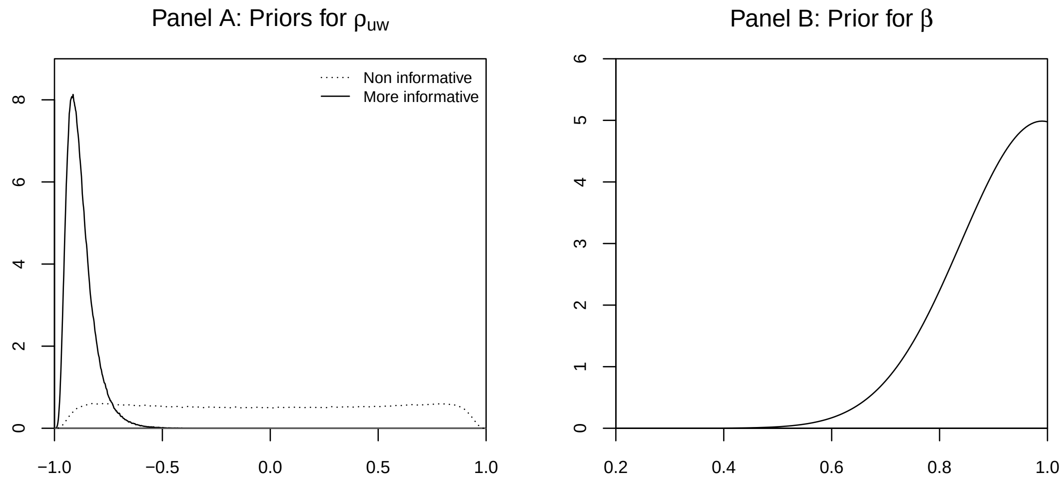

Figure 9: The prior distributions for ρ_{uw} and β used in the Bayesian analysis. Panel A plots the two prior distributions for ρ_{uw}: noninformative (flat between -0.9 and 0.9) and more informative (most of the mass below -0.71). Panel B plots the prior on the autoregressive coefficient β in the dynamics of μ.

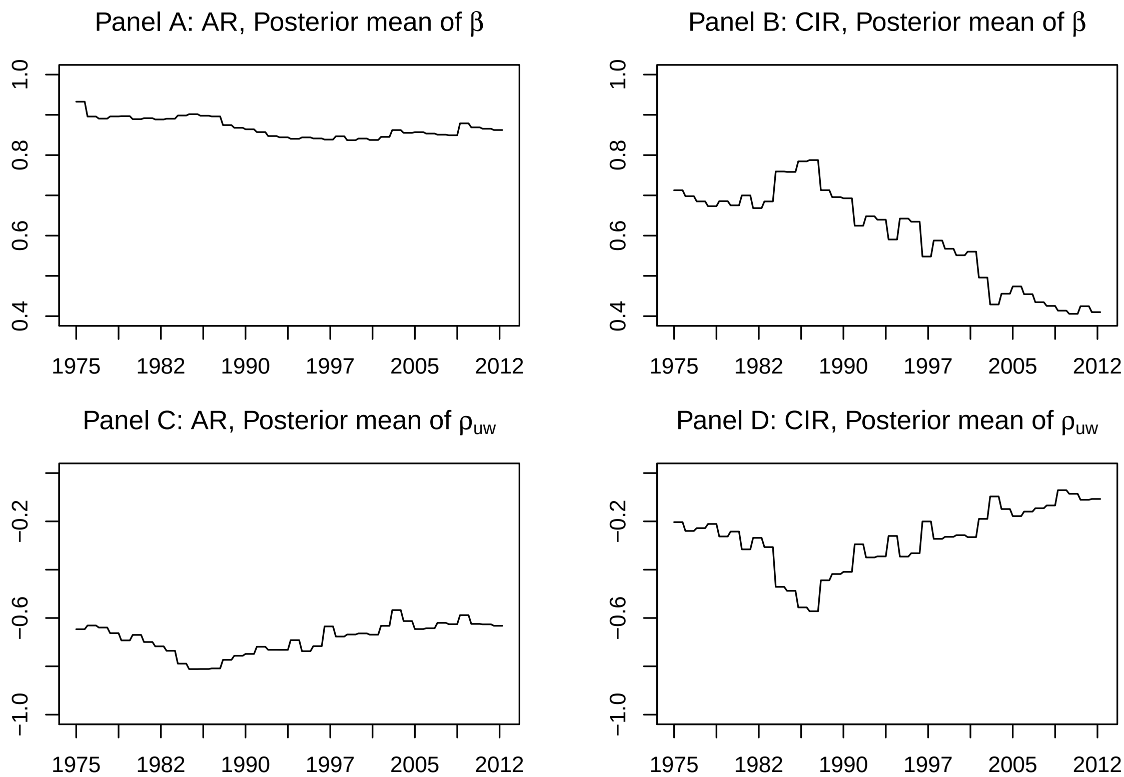

Figure 10: Evolution of the mean of the posterior distributions of β and ρ_{uw} for the AR and CIR predictive systems. The priors are more informative on ρ_{uw} and correspond to less predictability prior on R^2. The considered out-of-sample period is 1975Q1-2012Q4. Predictive systems are re-estimated on the first available date of each year in the sample.

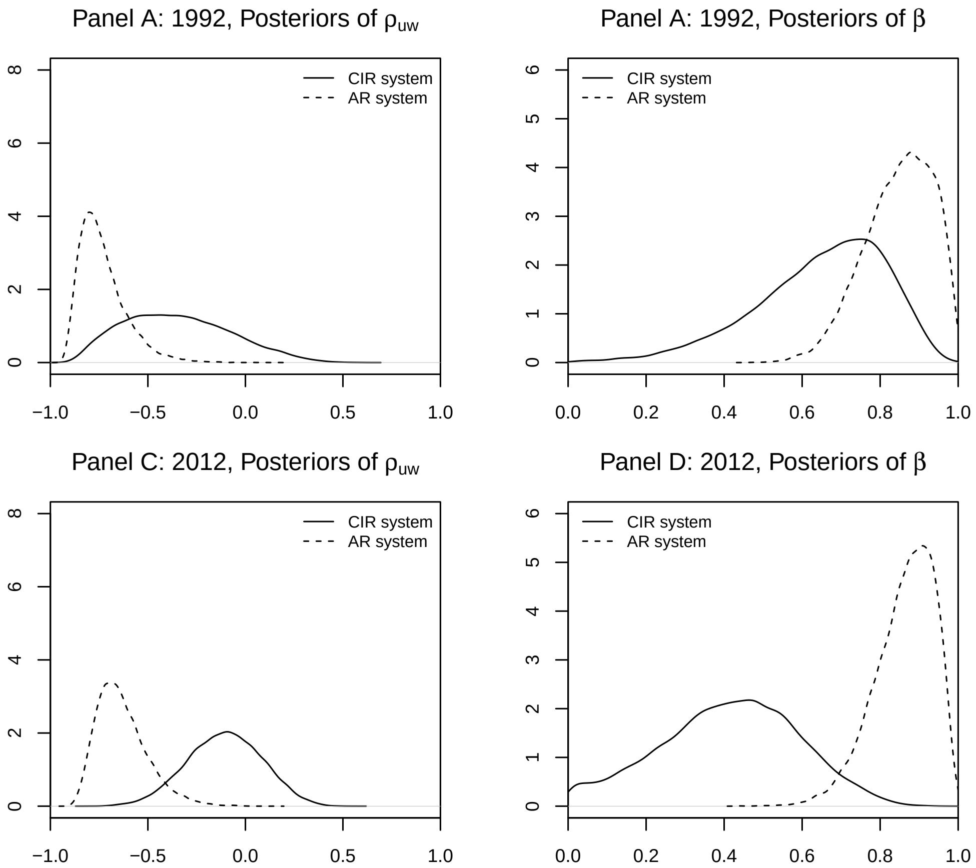

Figure 11: Posterior distributions of β and ρ_{uw} for the AR and CIR predictive systems. The priors are more informative on ρ_{uw} and correspond to less predictability prior on R^2. The posterior distributions are obtained from systems estimation in 1992-Q1 and 2012-Q1.

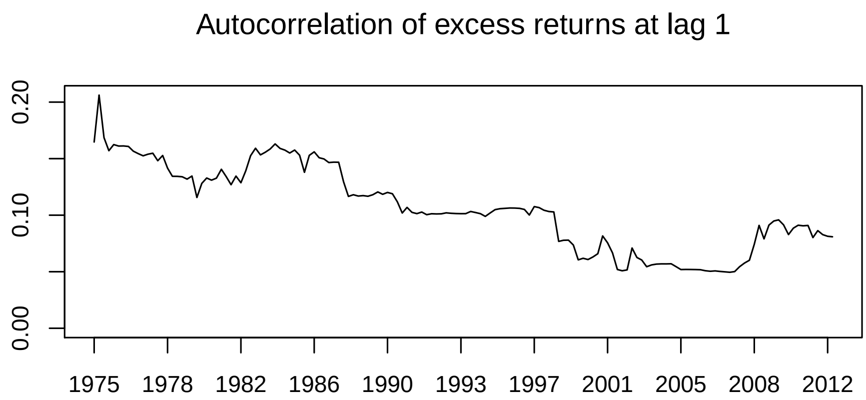

Figure 12: Evolution of the autocorrelation at lag 1 of returns. Our sample corresponds to quarterly returns on the value-weighted portfolio of all NYSE, Amex, and Nasdaq stocks in excess of the quarterly return on a 1-month T-bill obtained from CRSP. The sample begins in 1952. Each quarter, we re-estimate the autocorrelation at lag 1 using the available returns.

References

Ait-Sahalia, Y. and R. Kimmel (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics 83*(2), 413–452.

Alfonsi, A. (2005). On the discretization schemes for the cir (and bessel squared) processes. *Monte Carlo Methods and Applications 11*(4), 355.

Anderson, B. D. and J. B. Moore (2012). *Optimal filtering*. Courier Dover Publications.

Anderson, T. W. (2011). *The statistical analysis of time series*, Volume 19. John Wiley & Sons.

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

Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics 50*(4), 745–787.

Barreto, H. and F. Howland (2006). *Introductory econometrics: using Monte Carlo simulation with Microsoft excel*. Cambridge University Press.

Berkaoui, A., M. Bossy, A. Diop, et al. (2008). Euler scheme for sdes with non-lipschitz diffusion coefficient: strong convergence. *ESAIM: Probability and Statistics 12*.

Binsbergen, V., H. Jules, and R. S. Koijen (2010). Predictive regressions: A present-value approach. *The Journal of Finance 65*(4), 1439–1471.

Bossy, M., A. Diop, et al. (2007). An efficient discretisation scheme for one dimensional sdes with a diffusion coefficient function of the form| x|ˆ a, a in [1/2, 1).

Brandt, M. W. and Q. Kang (2004). On the relationship between the conditional mean and volatility of stock returns: A latent var approach. *Journal of Financial Economics 72*(2), 217–257.

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

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

Campbell, J. Y. (1991). A variance decomposition for stock returns. Technical report, National Bureau of Economic Research.

Campbell, J. Y. and J. Ammer (1993). What moves the stock and bond markets? a variance decomposition for long-term asset returns. *The Journal of Finance 48*(1), 3–37.

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

Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika 81*(3), 541–553.

Cochrane, J. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies 21*(4), 1533.

Cochrane, J. H. (2005). Time series for macroeconomics and finance. *Manuscript, University of Chicago*.

Connor, G. (1997). Sensible return forecasting for portfolio management. *Financial Analysts Journal 53*(5), 44–51.

Conrad, J. and G. Kaul (1988). Time-variation in expected returns. *Journal of Business*, 409–425.

Cox, J. C., J. E. Ingersoll Jr, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, 385–407.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*.

Duffee, G. R. (2007). Are variations in term premia related to the macroeconomy. *Unpublished working paper. University of California, Berkeley*.

Engle, R. (2001). Garch 101: The use of arch/garch models in applied econometrics. *Journal of Economic Perspectives*, 157–168.

Engle, R. F. and C. W. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251–276.

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

Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics 25*(1), 23–49.

Feller, W. (1951). Two singular diffusion problems. *Annals of mathematics*, 173–182.

Ferson, W. E. and C. R. Harvey (1991). The variation of economic risk premiums. *Journal of Political Economy*, 385–415.

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

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis 15*(2), 183–202.

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

Goyal, A. and I. Welch (2008). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *Review of Financial Studies 21*(4), 1455.

Hamilton, J. D. and G. Lin (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics 11*(5), 573–593.

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

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

Kim, C.-J. and C. R. Nelson (1999). State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books 1*.

Kvašňáková, K. (2013). Is imperfection better? evidence from predicting stock and bond returns. *Unpublished working paper. Vienna University of Economics and Business*.

Lamoureux, C. G. and G. Zhou (1996). Temporary components of stock returns: What do the data tell us? *Review of Financial Studies 9*(4), 1033–1059.

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

Lettau, M. and S. van Nieuwerburgh (2008). Reconciling the return predictability evidence. *Review of Financial Studies 21*(4), 1607.

Mantilla-Garcia, D. and V. Vaidyanathan (2011). Predicting Stock Returns in the presence of Uncertain Structural Breaks and Sample Noise. *Working Paper, Edhec Business School*.

McCracken, M. (2007). Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics 140*(2), 719–752.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics 8*(4), 323–361.

Merton, R. C. (1993). On the microeconomic theory of investment under uncertainty. *Handbook of Mathematical Economics 2*, 601–669.

Pages, G., F. Panloup, et al. (2009). Approximation of the distribution of a stationary markov process with application to option pricing. *Bernoulli 15*(1), 146.

Pástor, L. and R. Stambaugh (2009). Predictive systems: Living with imperfect predictors. *The Journal of Finance 64*(4), 1583–1628.

Pástor, L. and R. F. Stambaugh (2012). Are stocks really less volatile in the long run? *The Journal of Finance 67*(2), 431–478.

Pesaran, M. H. and A. Timmermann (1995). Predictability of stock returns: Robustness and economic significance. *The Journal of Finance 50*(4), 1201–1228.

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

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

Rousseeuw, P. J. and K. V. Driessen (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics 41*(3), 212–223.

Rytchkov, O. (2012). Filtering out expected dividends and expected returns. *The Quarterly Journal of Finance 2*(03).

Schwert, G. W. (1989). Why does stock market volatility change over time? *The Journal of Finance 44*(5), 1115–1153.

Stambaugh, R. F. (1986). Bias in Regressions with Lagged Stochastic Regressors. *Working Paper University of Chicago*.

Stambaugh, R. F. (1999). Predictive Regressions. *Journal of Financial Economics 54*, 375–421.

Van Binsbergen, J. H. and R. S. Koijen (2011). Likelihood-based estimation of exactly-solved present-value models. *Available at SSRN 1439849*.