City Research Online

City St George's, University of London

Citation: Sarno, L., Schneider, P. & Wagner, C. (2016). The Economic Value of Predicting Bond Risk Premia. Journal of Empirical Finance, 37, pp. 247-267. doi: 10.1016/j.jempfin.2016.02.001

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/13494/

Link to published version: https://doi.org/10.1016/j.jempfin.2016.02.001

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

The Economic Value of Predicting Bond Risk Premia*

Lucio Sarno† Paul Schneider‡ Christian Wagner§

February 10, 2016

Abstract

This paper studies whether the evident statistical predictability of bond risk premia translates into economic gains for investors. We propose a novel estimation strategy for affine term structure models that jointly fits yields and bond excess returns, thereby capturing predictive information otherwise hidden to standard estimations. The model predicts excess returns with high regressions R^2s and high forecast accuracy but cannot outperform the expectations hypothesis out-of-sample in terms of economic value, showing a general contrast between statistical and economic metrics of forecast evaluation. More specifically, the model mostly generates positive (negative) economic value during times of high (low) macroeconomic uncertainty. Overall, the expectation hypothesis remains a useful benchmark for investment decisions in bond markets, especially in low uncertainty states.

JEL classification: E43, G12.

Keywords: term structure of interest rates; expectations hypothesis; affine models; risk premia; statistical predictability; economic value.

*We are indebted to two anonymous referees, Geert Bekaert, Mike Chernov, Anna Cieslak, Greg Duffee, Bjørn Eraker, Alois Geyer, Amit Goyal, Hanno Lustig, Eberhard Mayerhofer, Antonio Mele, Dan Thornton, and Ilias Tsiakas for useful comments. The authors alone are responsible for any errors and for the views expressed in the paper.

†Cass Business School and Centre for Economic Policy Research (CEPR), London. Corresponding author. Faculty of Finance, Cass Business School, City University, London EC1Y 8TZ, UK. lucio.sarno@city.ac.uk.

‡Institute of Finance, University of Lugano, Via Buffi 13, CH-6900 Lugano. paul.schneider@usi.ch.

§Department of Finance, Copenhagen Business School, DK-2000 Frederiksberg, Denmark. cwa.fi@cbs.dk.

1 Introduction

Empirical research documents that the expectations hypothesis (EH) of the term structure of interest rates is rejected by the data and argues, almost unequivocally, that deviations from the EH reflect time-varying risk premia.[1] Fama (1984), Fama and Bliss (1987), and Campbell and Shiller (1991) are among the first to provide such evidence, while more recent studies that document the violation of the EH include Bekaert and Hodrick (2001) and Sarno et al. (2007). This evidence is strengthened by work showing that bond risk premia are predictable; see e.g. Cochrane and Piazzesi (2005). In this paper, we evaluate the relevance of EH deviations by studying whether bond investors benefit from conditioning on information about time-varying risk premia.

We estimate risk premia using affine term structure models (ATSMs). Based on the pioneering work of Duffie and Kan (1996) and Dai and Singleton (2000), ATSMs receive a particular focus in the finance literature on dynamic term structure models because of their richness, tractability, and ability to produce reasonable risk premium dynamics. Interestingly, research on the EH and on ATSMs has, to a large extent, evolved along separate paths.[2] Only a few papers attempt to bridge this gap and, for example, the results of Backus et al. (2001) and Dai and Singleton (2002) support the notion that the failure of the EH is due to the invalid assumption of constant risk premia. While Balduzzi and Chang (2012) find that ATSMs capture yield dynamics well, recent research argues that the evident predictability of bond risk premia cannot by captured by ATSMs because the necessary predictive information is not spanned

[1] The EH is the postulate that the long-term interest rate is determined by the current short-term rate and the market expectation of the short-term rate over the maturity of the long-term rate, plus a constant risk premium. Under the EH, pure discount bonds are perfect substitutes and excess returns are not predictable.

[2] While empirical EH research often argues that the theory's failure is due to time-varying risk premia, these papers put little effort into modeling risk premia, focusing instead on formal statistical tests of the EH. Similarly, research on ATSMs is usually motivated by the empirical rejection of the EH, but does not establish a direct link to the EH within the model.

by the cross-section of yields (see e.g. Duffee, 2011; Barillas, 2013; Joslin et al., 2013). By contrast, we show that such ATSMs do capture the predictability of excess returns when employing an extended estimation procedure that jointly fits yields and past risk premia to the data. This finding suggests that ATSMs represent a suitable vehicle for evaluating the economic consequences of EH deviations for bond investors.

Our paper contributes to the literature by evaluating whether ATSM forecasts are statistically more accurate and economically more valuable than EH-consistent forecasts or, in contrast, whether presuming that the EH holds is a suitable first-order approximation for bond investment decisions. We conduct an empirical evaluation of the EH that is in many respects more comprehensive than evaluations in previous research. First, using ATSMs, we consistently model the whole term structure and not only a subset of yields or excess returns, as e.g. in Fama and Bliss (1987), Campbell and Shiller (1991), Bekaert and Hodrick (2001), Cochrane and Piazzesi (2005). Second, the extended estimation proposed in this paper accounts for predictive information in and beyond (i.e. unspanned by) the term structure, thereby producing a stronger challenge to the EH.[3] Third, while related research generally either focuses on a particular segment of the term structure or a specific prediction horizon, we analyze the term structure of bond risk premia for horizons ranging from one month to five years. Fourth, while many other papers focus on statistical evidence in-sample – e.g. Fama and Bliss (1987), Campbell and Shiller (1991), Bekaert and Hodrick (2001), Cochrane and Piazzesi (2005) – we measure both the statistical accuracy as well as the economic value added by conditional risk premia (Leitch and Tanner, 1991), and we complement the in-sample results with an out-of-sample analysis. Our paper is thus related to, but more general than Thornton and Valente (2012),

[3]Recent research suggests that such additional information that adds to statistical predictability may originate from forward rates (Cochrane and Piazzesi, 2005), macroeconomic factors (see, e.g., Ludvigson and Ng, 2009; Cieslak and Povala, 2013; Joslin et al., 2013), the market variance risk premium (Mueller et al., 2011), or a 'hidden factor' (Duffee, 2011).

who specifically investigate the economic value of one-year out-of-sample forecasts using the Cochrane and Piazzesi (2005) factor as well as compared to Barillas (2013) focusing on the role of macroeconomic variables. Finally, our paper is related to Adrian et al. (2013), who propose a regression-based ATSM estimation that incorporates information about bond excess returns. By construction, their approach fits realized bond excess returns almost perfectly. However, we show that this framework generates predictions for bond risk premia that display the same tension between statistical and economic metrics as our model, thus supporting the general findings reported using our extended estimation.

Using US yield data from 1952 to 2012, we evaluate 25 combinations of prediction horizons and bond maturities, with maturities ranging from one month to ten years. We find that the extended estimation increases predictive ability and adds economic value over the standard estimation used in the literature. Conditional risk premia from the extended estimation are unbiased, thereby explaining deviations from the EH, and entail high explanatory power for bond excess returns, beyond R^2s reported in related work. These findings suggest that our estimation strategy is flexible enough to capture long- and short-term predictive information from different sources. As a result, the model allows bond investors to forecast risk premia with high accuracy and to earn positive portfolio excess returns in- and out-of-sample.

To evaluate the model against the EH, we use the average historical bond excess return as a consistent estimate for the EH-postulate of constant risk premia. The extended estimation beats the EH in terms of statistical forecast accuracy but the model's predictive ability does not lead to superior portfolio performance out-of-sample relative to the EH. Thus, the EH still provides a useful out-of-sample benchmark, and we view the finding that bond investors generally cannot benefit from using conditional risk premia relative to using the historical average as the bond market analogue to the result of Goyal and Welch (2008) for stock markets.

We also provide a general discussion on why conclusions based on statistical metrics of forecast accuracy may deviate from those reached using economic value measures. On the one hand, EH deviations may be statistically significant, but too small to be meaningfully exploited by bond investors. On the other hand, common predictive ability measures evaluate loss functions that are in many respects unrelated to the economic success of bond investments. As a consequence it cannot be taken for granted that even models with high forecast accuracy allow for economically meaningful bond investment returns. We illustrate the validity of these general arguments using the results of our model estimations, but these arguments are equally valid for the mounting number of papers on statistical predictability of bond excess returns.

Finally, on a more positive note, we find that our proposed ATSM performs better than the EH in terms of economic value during periods of high macroeconomic uncertainty. This result is intriguing and has a natural interpretation: during times of low uncertainty risk premia are fairly constant and hence the EH provides a very hard benchmark to beat, whereas during high uncertainty periods bond risk premia are more volatile and hence a rich ATSM that explicitly models the time variation in risk premia adds economic value.

We discuss the model specification and estimation in Section 2. Section 3 details the data and reports descriptive statistics for yield pricing errors. We present empirical results on the evaluation of the extended versus the standard estimation procedure and relative to the EH in Section 4. Section 5 contains a discussion of our results and Section 6 concludes. The separate Internet Appendix contains technical details and additional empirical results.

2 Empirical Model and Estimation

Consider a long-term bond with T years maturity and a short-term bond with τ years maturity. We denote by p_t^T the time-t price of a T-year zero coupon bond with a certain payoff of 1 at maturity. The corresponding (effective) yield is given by

$$y_t^T = -\log[p_t^T]. \tag{1}$$

Analogously, we use the notation p_t^τ and y_t^τ for the price and the yield of the short-term bond with $\tau \leq T$. The prices of the short- and long-term bonds imply the time-t forward rate effective for $T-\tau$ periods beginning at $t+\tau$

$$f_{t,\tau}^{T-\tau} = \log[p_t^\tau / p_t^T]. \tag{2}$$

The return of buying a T-year bond at time t and selling it at time $t+\tau$ ($h_{t+\tau}^T$) is given by $h_{t+\tau}^T = \log[p_{t+\tau}^{T-\tau}/p_t^T]$, and the corresponding bond excess return ($rx_{t+\tau}^T$) is thus

$$rx_{t+\tau}^T \equiv f_{t,\tau}^{T-\tau} - y_{t+\tau}^{T-\tau}. \tag{3}$$

The EH presumes that the forward rate is equal to the expected yield (under the physical probability measure) plus a constant risk premium. To accommodate potentially time-varying risk premia, we now turn to the specification of an affine term structure model (ATSM).

2.1 Affine Term Structure Model and Bond Risk Premia

Based on the findings of Litterman and Scheinkman (1991), it has become well-established practice to employ term structure models with three factors. Accordingly, we use an ATSM with three factors, which we specify as a purely Gaussian $A_0(3)$ model. We consider two parameterizations of the model. First, we estimate the ATSM under latent state variables. Second, we use the observable states parametrization of Joslin et al. (2011, JSZ). We describe bond pricing and risk premia in terms of a generic affine model below but delegate the detailed specifications of the latent and observable factors models to Appendix A.

2.1.1 Affine Term Structure Model

For our empirical analysis, we use a continuous-time affine term structure model for an economy that is driven by state variables X living on a canonical state space $\mathcal{D} = \mathbb{R}^m_+ \times \mathbb{R}^n, m, n \geq 0$, $d = m + n \geq 1$. Under a given probability measure $\mathbb{M}$ the evolution of X solves the stochastic differential equation

$$dX_t = (b^{\mathbb{M}} - \beta^{\mathbb{M}} X_t)dt + \sigma(X_t)dW_t^{\mathbb{M}}, \tag{4}$$

where $\sigma(x)\sigma(x)^\top = a + \alpha x$, a is a $d \times d$ matrix, and α is a $d \times d \times d$ cube. Throughout we assume boundary non-attainment conditions for $X_{i,t}, 1 \leq i \leq m$ in order to ensure existence of transition densities (Filipović et al., 2013) and to use generalized affine market prices of risk from Cheridito et al. (2007) in addition to the admissibility conditions from Duffie et al. (2003). This means that $2b_i^{\mathbb{M}} > \alpha_{i,ii}, 1 \leq i \leq m$. In what follows we will make use of two specific probability measures: $\mathbb{Q}$, the pricing measure, and $\mathbb{P}$, the time-series measure. To avoid overfitting and to reduce estimation noise we impose a lower-triangular form of the mean-reversion matrix $\beta^{\mathbb{M}}$ for $\mathbb{M} \in \{\mathbb{P}, \mathbb{Q}\}$. Furthermore, we restrict its diagonal to strictly positive values. This ensures

a stationary system and existence of unconditional moments. The remaining parameterization (in particular the diffusion function) is modeled in its most flexible form according to the Dai and Singleton (2000) specification, respectively with lower-triangular diffusion function for the observable JSZ model.

The instantaneous short rate is affine in X, $r(t) \equiv \delta_0 + \delta_X^\top X_t$, which implies that bond prices p_t^T are exponentially affine in the state variables X

$$p_t^T = \mathbb{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{t+T} r(u)\, du} \right] = e^{\phi(T) + \psi(T)^\top X_t}, \tag{5}$$

where ϕ and ψ solve the ordinary differential equations

$$\dot{\psi} = -\delta_X - \beta^{\mathbb{Q}\top}\psi + \frac{1}{2}\psi^\top \alpha \psi, \qquad \psi(0) = 0, \tag{6}$$

$$\dot{\phi} = -\delta_0 + b^{\mathbb{Q}\top}\psi + \frac{1}{2}\psi^\top a \psi, \qquad \phi(0) = 0. \tag{7}$$

We collect the set of parameters governing the evolution of X by defining $\theta^{\mathbb{P}} \equiv \left\{ b^{\mathbb{P}}, \beta^{\mathbb{P}}, a, \alpha \right\}$, $\theta^{\mathbb{Q}} \equiv \left\{ b^{\mathbb{Q}}, \beta^{\mathbb{Q}}, a, \alpha, \delta_0, \delta_X \right\}$, and $\theta^{\mathbb{Q}\mathbb{P}} \equiv \theta^{\mathbb{Q}} \cup \theta^{\mathbb{P}}$. The coefficients ψ and ϕ are functions of time and the parameters, but we will suppress this dependence if the context permits.

2.1.2 Bond Risk Premia: Conditional Expectations of Bond Excess Returns

We combine Eq. (5) with Eqs. (1) and (2) to express the yield and the forward rate as

$$y_t^T = -\log[p_t^T] = -(\phi(T) + \psi(T)^\top X_t), \tag{8}$$

$$f_{t,\tau}^{T-\tau} = \phi(\tau) - \phi(T) + (\psi(\tau) - \psi(T))^\top X_t. \tag{9}$$

Using these relations, we calculate expected yields and expected excess returns. To appreciate the structure of the risk premium induced through the affine state variables we note that

$$\mathbb{E}_t^{\mathbb{P}}\left[X_{t+\tau}\right] = A(\tau) + B(\tau)X_t, \tag{10}$$

where $B(\tau) = e^{-\beta^{\mathbb{P}}\tau}$ and $A(\tau) = b^{\mathbb{P}} \int_0^\tau B(u)du$. We then express conditional expectations as

$$\mathbb{E}_t^{\mathbb{P}}\left[y_{t+\tau}^{T-\tau}\right] = -(\phi(T-\tau) + \psi(T-\tau)^\top(A(\tau) + B(\tau)X_t)), \tag{11}$$

$$\mathbb{E}_t^{\mathbb{P}}\left[rx_{t,\tau}^{T}\right] = (\psi(\tau) - \psi(T))^\top X_t + \psi(T-\tau)^\top(A(\tau) + B(\tau)X_t). \tag{12}$$

Making explicit the dependence of ϕ and ψ on the parameters and introducing

$$\gamma^{\tau,T}(\theta_{\mathbb{QP}}) \equiv \psi(\tau,\theta_{\mathbb{Q}})^\top - \psi(T,\theta_{\mathbb{Q}})^\top + \psi(T-\tau,\theta_{\mathbb{Q}})^\top B(\tau,\theta_{\mathbb{P}}), \tag{13}$$

$$\eta^{\tau,T}(\theta_{\mathbb{QP}}) \equiv \psi(T-\tau,\theta_{\mathbb{Q}})^\top A(\tau,\theta_{\mathbb{P}}), \tag{14}$$

the time-t risk premium is affine in η and γ

$$\mathbb{E}_t^{\mathbb{P}}\left[rx_{t,\tau}^{T}\right] = \eta^{\tau,T} + \gamma^{\tau,T}X_t. \tag{15}$$

The risk premium in Eq. (15) depends on τ, T, and on t (through X). It comprises a constant as well as a time-varying component that is driven by the evolution of X_t, which can be seen from rewriting the conditional expectation in Eq. (15) as

$$\mathbb{E}_t^{\mathbb{P}}\left[rx_{t,\tau}^{T}\right] = \eta^{\tau,T} + \gamma^{\tau,T}\mathbb{E}^{\mathbb{P}}\left[X\right] - \gamma^{\tau,T}\left(\mathbb{E}^{\mathbb{P}}\left[X\right] - X_t\right). \tag{16}$$

This relation interprets the time-variation in risk premia as deviations of X_t from its unconditional expectation. The first two terms only depend on τ and T and are thus time-invariant, consistent with the EH notion of a constant risk premium. Empirically, the question whether the EH holds can be assessed by analyzing whether the last term, which should be just noise under the EH, induces predictability of bond excess returns. Note that when estimating the model, the sum of the first two terms will correspond to the average excess return observed in the data and the last term will average to zero. In that sense, the time-invariant part determines for a given horizon the shape of the (average) term structure of risk premia. Building on these insights from Eqs. (15) and (16), we estimate the EH-postulated constant risk premia using historical sample averages of bond excess returns. To estimate ATSM-implied conditional risk premia that additionally capture the time-varying component, we employ the estimation methodology described in the next section.

2.2 Model Estimation

For our empirical analysis we distinguish between the model specification in terms of latent factors and the JSZ specification in terms of observable factors. We follow two different estimation strategies for each of the two specifications. The first, standard estimation procedure, requires model-implied yields to match the observed term structure. The second, extended estimation, requires that additionally model-implied bond excess returns match past realized excess returns. To accommodate the notion of an investor updating her beliefs about the model's predictability and to include past failures and successes into estimates of the parameters (and state variables), we employ Bayesian methodology for the latent-factor model and maximum likelihood (as well as nonlinear least squares) for the JSZ specification. Without changing the structure of the model, this explicitly accounts for the time-series properties of EH deviations

in addition to the cross-sectional properties of yields. With this novel approach we account for information that is not embedded in the term structure of interest rates but adds to predictive ability for bond excess returns.

We assess the forecast performance of ATSMs estimated with the standard and with the extended procedure, both in-sample and out-of-sample. For the in-sample analysis, we estimate ATSMs using the full set of data available. In the out-of-sample analysis, we generate conditional time-t expectations by estimating ATSMs using only information that is available at time t. We first estimate the models using the earliest 120 months of data available. Subsequently, we update the information set every month and re-estimate the models (parameters and state variables) to generate updated out-of-sample forecasts. We present a concise description of the two estimation strategies below and provide technical details of the Markov-Chain Monte Carlo (MCMC) methods that we apply for the standard and extended estimation procedures in Appendix A.1. The estimation of the JSZ specification is outlined in Appendix A.2.

2.2.1 Standard Estimation Procedure

Our data set comprises zero yields with 24 maturities (expressed in years) $T_1, \ldots, T_{24}$, covering 1, 2, 3, 4, 6, 7, 9, 12 ,13, 15, 18, 24, 25, 27, 30, 36, 48, 60, 61, 63, 66, 72, 84, 120 months; for details about the data, see Section 3. We estimate our model using observation equations

$$\frac{y_t^{T_i}}{T_i} = -\frac{\phi(T_i, \theta^{\mathbb{Q}}) + \psi(T_i, \theta^{\mathbb{Q}})^\top X_t}{T_i} + \varepsilon_t^{T_i}, \tag{17}$$

where $\varepsilon_t^{T_i}, i = 1, \ldots, 24$ are assumed i.i.d normally distributed with mean zero and $\mathbb{V}\left[\varepsilon_t^{T_i}\right] = e^{-2\left(c_0 + c_1 T_i + c_2 T_i^2\right)}$. For the latent-state model we use these equations for filtering and smoothing the state variables X and define $\theta^\epsilon \equiv \{c_0, c_1, c_2\}$ and finally $\theta \equiv \theta^{\mathbb{QP}} \cup \theta^\epsilon$. The JSZ model uses

this equation for determining the fit to the yield curve for fixed observed state variables.

For the latent-factor model in a Bayesian setting, with discretely observed data sample at times $t_1, \ldots, t_N$ the joint log posterior ℓ of the latent states with the parameters for a window $[t_m, t_n], t_1 \leq t_m < t_n \leq t_N$ is

$$\ell_m^n(\theta, X) = \sum_{k=m}^{n} \Big\{ \log p(X_{t_k} \mid X_{t_{k-1}}, \theta^{\mathbb{P}}) + \sum_{i=1}^{24} \log p(\varepsilon_{t_k}^{T_i} \mid \theta_\epsilon) + \log \pi(\theta) \Big\}, \tag{18}$$

with the prior

$$\pi(\theta_i) \propto \begin{cases} \mathbb{1}_{\{\theta_i \text{ admissible}\}} & \theta_i \in \mathbb{R} \\ \frac{\mathbb{1}_{\{\theta_i \text{ admissible}\}}}{\theta_i} & \theta_i \in \mathbb{R}_+ \end{cases}. \tag{19}$$

The first term on the right hand side of Eq. (18) contains the transition densities, the second reflects yields pricing errors, and the third the prior distribution of the parameters. Draws θ, X from the complicated distribution in Eq. (18) are obtained by sampling in turn from $X \mid \theta$ and $\theta \mid X$. The JSZ estimation uses the same equation without the prior density. For the in-sample analysis, we estimate the ATSM once using the full data set, i.e. using $[t_1, t_N]$. In the out-of-sample analysis, we start by estimating the model for the first 120 months of data available ($[t_1, t_{120}]$) to generate forecasts of τ-month excess returns to be realized at $t_{120+\tau}$. Subsequently, we re-estimate the model every month using the expanded information set to generate new forecasts; i.e. at time t_{120+j} we estimate the model using the window $[t_1, t_{120+j}]$ to generate forecasts of τ-period bond excess returns realized at time $t_{120+j+\tau}$.

2.2.2 Extended Estimation Procedure

Bond investors pay close attention to bond excess returns and evaluate past forecast errors to account for this information in their predictions and portfolio choices. To reflect this behavior

we propose an extended estimation which matches model risk premia with past realized excess returns using Eq. (15). We therefore additionally consider the set of all possible (34) forecast equations given the available yield maturities

$$f_{t,\tau_i}^{T_{i,j}-\tau_i} - y_{t+\tau_i}^{T_{i,j}-\tau_i} = \eta^{\tau_i,T_{i,j}}(\theta^{\mathbb{QP}}) + \gamma^{\tau_i,T_{i,j}}(\theta^{\mathbb{QP}})X_t + \epsilon_{t+\tau_i}^{\tau_i,T_{i,j}}. \tag{20}$$

The forecast errors $\epsilon_{t+\tau_i}^{\tau_i,T_{i,j}}$ are assumed i.i.d normal with mean zero and variance $\mathbb{V}\left[\epsilon_{t+\tau_i}^{\tau_i,T_{i,j}}\right] = e^{-2\left(D\left(c_0+c_1T_{i,j}+c_2T_{i,j}^2\right)+\left(d_0+d_1\tau_i+d_2\tau_i^2\right)\right)}$. We now define $\theta^{\epsilon\varepsilon} \equiv \{c_0, c_1, c_2, d_0, d_1, d_2, D\}$ and finally $\theta \equiv \theta^{\mathbb{QP}} \cup \theta^{\epsilon\varepsilon}$, and use Eq. (20) in addition to Eq. (17) for filtering and smoothing the latent state variables X. The joint, augmented log posterior $\tilde{\ell}$ of the latent states with the parameters is now[4]

$$\begin{aligned}\tilde{\ell}_m^n(\theta, X) = \sum_{k=m}^{n} \Bigg\{ &\log p(X_{t_k} \mid X_{t_{k-1}}, \theta^{\mathbb{P}}) + \sum_{i=1}^{24} \log p(\varepsilon_{t_k}^{T_i} \mid \theta_{\epsilon\varepsilon}) \\ &+ \sum_{1\le i\le 5, 1\le j\le J_i} \log p(\epsilon_{t_k}^{\tau_i,T_{i,j}} \mid \theta_{\epsilon\varepsilon}) \mathbb{1}_{\{t_k+\tau_i\le t_n\}} \Bigg\} + \log \pi(\theta),\end{aligned} \tag{21}$$

with $\pi(\theta_i)$ as in Eq. (19). The first term in the second line of Eq. (21) reflects the excess return forecast errors ε, which affect estimates of θ and X.[5,6] As before in the standard estimation, the JSZ model uses the same likelihood equation without the prior density.

Similar to the standard estimation procedure, we use the window $[t_1, t_N]$ to estimate the ATSM for the in-sample analysis. To generate out-of-sample forecasts, we first estimate the

[4]The augmented likelihood contains a filtering (first line, second term) and a forecasting (second line, first term) component. The filtering component is necessary for out-of-sample forecasting. At time t_i the investor learns about realizations of latent states only from the time-t_i term structure and makes her forecast.

[5]Augmenting the likelihood with forecast errors, any information in bond excess returns is absorbed by the latent states and parameters regardless of the drivers. If the data were truly Markovian, the forecast equations would be irrelevant and not affect parameter and state variable estimates. We allow past forecast errors to affect state variable and parameter estimates to admit a *learning effect*, but we do not build learning into conditional expectations directly, a computationally intensive approach taken by Barberis (2000).

[6]Note from Eq. (20) that the procedure of matching risk premia also incorporates information from forward rates, which Cochrane and Piazzesi (2005) find to be an important source of predictability.

model on the earliest 10 years of data available (i.e. window $[t_1, t_{120}]$) and then re-estimate the model every month using the expanded information set to generate new forecasts (i.e. window $[t_1, t_{120+j}]$). We stress here that in the out-of-sample bond investment decision to be made at time t_{120+j}, the investor first samples from the joint distribution of the parameters and latent states through the augmented likelihood Eq. (21) using only forecast error information available up to time t_{120+j}. For each draw of θ and X from this joint distribution she then makes an out-of-sample forecast, records it, and with enough draws (we use 100,000) chooses the sample mean of all recorded forecasts as the forecast to be used in her investment decision. The JSZ model performs the same step, but keeping the (observable) state variables fixed.

3 Data and Yield Pricing Errors

We construct a data set of monthly US interest rates with maturities ranging from one month to ten years from 1952 to 2012. For the period up to 2003, we use the yield data of Sarno et al. (2007), which is virtually identical to that of Campbell and Shiller (1991) over the respective period (1952 to 1987). For the period 2004 to 2012, we obtain short-term yields (maturities less than one year) from the CRSP Fama T-Billl Structure and long-term yields from the Treasury curve published by the Fed (Gürkaynak et al., 2007). Our results are thus directly comparable to the large EH and bond risk premium literature on the US market.

Using this data, we estimate the latent factor and the observable factor models described in Section 2. We delegate detailed estimation results (including parameter estimates as well as rotation and interpretation of state variables) to Appendix B, because they are not crucial for our main objective. Table 1 summarizes the models' yield pricing accuracy when using the standard estimation and the extended estimation procedure that also matches risk premia.

For the latent factor model, the standard estimation fits yields better, with root mean squared errors (RMSEs) and standard deviations of pricing errors across maturities of 17 basis points as compared to 26 basis points for the extended estimation. These magnitudes are comparable to numbers reported in related research and suggest that both estimation strategies match the term structure of yields satisfactorily. For the observable factor model, the pricing errors are even lower with RMSEs of around 8 basis points across maturities for both estimation procedures. The difference in yield pricing errors for the latent and observable factor models is mostly driven by the latter fitting short-term yields more accurately. This is a consequence of the JSZ assumption that a linear combination of yields, the principal components, is observed without error. Since the first principal component is strongly related to the level of the yield curve, short rates are fitted tighter by construction.

4 Forecasting Bond Excess Returns and Economic Value

We now evaluate the statistical accuracy and economic value of bond excess return forecasts generated by latent and observable factor ATSMs in- and out-of-sample. We document that investors are willing to pay a sizable premium to switch from the standard to the extended estimation; however, investors cannot systematically benefit out-of-sample compared to using forecasts of EH-postulated constant risk premia, which we consistently estimate as averages of historical bond excess returns.

4.1 Bond Risk Premium Regressions

Table 2 presents results for regressing realized bond excess returns on model risk premia for 25 combinations of horizons and maturities. We assess the significance of the slope coefficients b

by calculating standard errors following Hansen and Hodrick (1980).

We start by presenting results for the latent factor model in Panel A. For the standard estimation more than half of the slope estimates are significantly positive and many are close to one. Across longer-term bond maturities, the average one-month and one-year prediction horizon R^2s are 1% and 19%, respectively. Model-implied risk premia from the extended estimation are generally significant and unbiased predictors of realized excess returns that have high explanatory power with R^2s of around 11% and 87% over one-month and one-year prediction horizons. For the observable factor model (Panel B) there is little difference between the standard and extended estimation regressions results. The slope estimates are significant, however, the explanatory power relative to the latent factor model is low, with cross-maturity R^2s being 2% and 18% at the one-month and one-year horizons.

Comparing the extended estimation results of the latent and observable factor models suggests that there is a trade-off between fitting yields (Table 1) and fitting bond excess returns: The latent factor model matches bond risk premia at the expense of higher yield pricing errors while the opposite is true for the observable factors model. For the latent factor model, the extended model estimation clearly dominates the standard estimation in terms of explanatory power for realized excess returns. These results are consistent with previous research documenting that bond excess returns are predictable at shorter and longer horizons (see e.g. Cochrane and Piazzesi, 2005; Ludvigson and Ng, 2009; Cieslak and Povala, 2013; Mueller et al., 2011) and that this predictability is to a large extent not spanned by the term structure of bond yields and thus not captured in standard ATSM estimations (see e.g. Duffee, 2011). Finding that model expectations are unbiased supports the argument that accounting for risk premia can explain classical EH tests suggest a rejection of the EH (e.g. Dai and Singleton, 2002). In what follows, we take a closer look at the improvement in forecast accuracy and the economic

value that accrues to investors using the extended instead of the standard estimation procedure and relative to EH-consistent constant risk premium forecasts.

4.2 Statistical Accuracy of Bond Excess Return Forecasts

To evaluate the accuracy of extended estimation forecasts against the standard estimation and the EH constant risk premium benchmarks, we report values for a $R2$-metric defined similarly to Campbell and Thompson (2008)

$$R2 \equiv 1 - MSE^m/MSE^b, \tag{22}$$

where $MSE^k = 1/(N-\tau+1)\sum_{t=1}^{N-\tau}(rx_{t+\tau}^T - \mathbb{E}_t^{\mathbb{P},k}\left[rx_{t+\tau}^T\right])^2$ denotes the mean squared forecast error of the model ($k=m$) and the benchmark ($k=b$), respectively. $R2$ takes positive values when forecasts from model m are more accurate than those from benchmark model b and negative values when the opposite is the case.[7] To judge the significance of $R2$-statistics, we estimate confidence intervals as the 5%- and 95%-percentiles using a block bootstrap.[8]

For the latent factor model, Panel A of Table 3 shows that the in-sample $R2$ estimates of the extended versus the standard estimation are positive for all 25 horizon/maturity combinations with estimates being significant in most cases (as indicated by *). The extended estimation forecasts are also more accurate than constant risk premium forecasts with all $R2$ estimates being positive and statistically significant in 23 of 25 combinations. Out-of-sample, the extended outperforms the standard estimation, with positive (and significant) $R2$s in 15 (10) of 25 horizon/maturity combinations whereas only one $R2$ estimate is significantly negative. Using the

[7]Note that many common measures of predictive ability are based on squared loss functions (e.g. Diebold and Mariano, 1995) and therefore lead to the same conclusions that we reach in this paper using $R2$.

[8]To determine the optimal block size, we follow Politis and White (2004) and Patton et al. (2009). Results are qualitatively the same when using the simpler rules suggested in Hall et al. (1995).

EH benchmark, the latent model beats the EH at horizons of one and two years but not at shorter horizons, and for long-term horizons the results are mixed.

The observable factors model (Panel B) generates very similar in-sample predictive accuracy for the standard and extended estimation (in line with the regression results reported above). Out-of-sample, the extended estimation dominates the standard estimation with all $R2$ estimates being positive and, for 23 of 25 estimates, significant. Furthermore, the observable factor model outperforms the EH-implied constant risk premia in- and out-of-sample.

Overall, the extended estimation picks up information relevant for predicting bond risk premia that is hidden to affine models that are estimated by only fitting yields. Accounting for the information in forward rates and past bond excess returns improves the model's forecast accuracy in- and out-of-sample. Moreover, model forecasts are more accurate than constant risk premium forecasts in-sample. Out-of-sample, the results appear to be horizon-dependent for the latent factor model whereas the observable factor model beats the EH.

4.3 Economic Value of Bond Excess Return Forecasts

We now investigate whether superior predictive ability of the extended estimation compared to benchmark forecasts translates into economic benefits for bond investors. First, we evaluate optimal bond portfolios in the quadratic utility framework of West et al. (1993).[9] For investment horizon τ, the investor chooses to allocate his wealth between bonds with maturities τ and $T > \tau$. Since the maturity of the shorter-term bond matches the investment horizon, the τ-bond represents the risk-free asset. The longer-term bond, with remaining maturity $T - \tau$ at the end of the horizon, is the risky asset. Let $\mu_{t+\tau}^{T,k}$ denote the $N \times 1$ vector of conditional expectations of risky asset returns generated by model k and denote the associated covariance

[9]Della Corte et al. (2008) and Thornton and Valente (2012) also use this approach for US bond markets.

matrix by $\Sigma_{t+\tau}$.[10] For a given target volatility σ^*, we maximize the portfolio excess return to obtain the mean-variance optimal weights for the risky asset

$$w_t^k = \frac{\sigma^*}{\sqrt{C_t}} \Sigma_{t+\tau}^{-1} \mu_{t+\tau}^{T,k},$$

where w_t^k is a $N \times 1$ vector and $C_t = \mu_{t+\tau}^{T,k\top} \Sigma_{t+\tau}^{-1} \mu_{t+\tau}^{T,k}$. The weights of the riskless asset are given by $\mathbf{1} - w_t^k$, where $\mathbf{1}$ is a $N \times 1$ vector of ones and the resulting gross portfolio return from t to $t+\tau$ is given by $R_{t+\tau}^k = 1 + y_t^\tau + w_t^k \cdot rx_{t+\tau}^T$.

Second, we consider an investor with power utility and constant relative risk aversion ρ, so that the utility function is $U(W_{t+\tau}) = \frac{W_{t+\tau}^{1-\rho}}{1-\rho}$, where wealth $W_{t+\tau}$ is determined by initial wealth at time t and the performance of a portfolio containing a riskless bond and a risky bond with conditional variance σ_t^2. The optimal weight of the risky asset is given by

$$w_t^k = \frac{\mu_{t+\tau}^{T,k} - y_t^\tau + \frac{1}{2}\sigma_t^2}{\rho \sigma_t^2} \tag{23}$$

and the weight of the riskless asset is therefore $1 - w_t^k$. We provide a detailed description of the power utility framework in Appendix C.

To measure the economic value generated by model m over model b, we compute the performance measure Θ proposed by Goetzmann et al. (2007). Θ quantifies the risk-adjusted

[10] Analogous to the estimation of constant risk premia, we estimate covariances based on sample standard deviations of bond excess returns and based on ten-year expanding windows for the in-sample and out-of-sample analysis, respectively. We choose this simple approach to estimate covariances because the focus of the current paper is set on the predictability of the first moment of the bond excess return distribution and the term structure of bond risk premia. We repeat the economic value analysis also using covariance matrices estimated with ATSMs and find that there is no impact on our conclusions: (i) when we compare portfolio allocations based on standard estimation to extended estimation forecasts using the model-implied covariances, we find that switching to the extended estimation adds economic value irrespective of the ATSM specification considered; (ii) when we compare ATSM portfolios to constant risk premium portfolios, we find that returns of these portfolios may be somewhat different because the volatility level implied by the model estimates is different, but when we consider the risk-adjusted measures described in this section, our conclusions remain unchanged. Furthermore, simple linear regression and rolling sample variance estimates, as for instance in Thornton and Valente (2012), lead to the same conclusions.

premium return that the portfolio based on forecasts from model m earns in excess of the benchmark portfolio and is calculated as

$$\Theta = \frac{12}{(1-\rho)\tau} \ln \left(\frac{1}{N-\tau+1} \sum_{t=1}^{N-\tau} [(1+R_n^m)/(1+R_n^b)]^{1-\rho} \right). \tag{24}$$

In contrast to the commonly reported Sharpe ratio, Θ alleviates concerns related to non-normality. Furthermore, compared to the performance fee of Fleming et al. (2001) it does not assume a specific utility function.[11] Throughout the empirical analysis we set $\sigma^* = 2\%$ *p.a.*, $\rho = 3$, and impose a maximum leverage of 100%; all our results are robust to choosing other values.

We report portfolio excess returns of investors using forecasts from the extended estimation and performance measures relative to the standard estimation and EH forecasts in Tables 4 and 5 for mean-variance and power utility investors, respectively. Mean-variance investors earn positive portfolio excess returns that tend to increase with the maturity of the longer-term bond and decrease with prediction horizon. For the latent factor model (Panel A), the extended estimation dominates the standard estimation by generating Θ values that are positive for all horizon/maturity combinations in-sample and in 23 of 25 combinations out-of-sample. Premium returns in excess of EH portfolios are also positive in 21 of 25 scenarios in-sample. Θ estimates increase with bond maturity but decrease with prediction horizon, suggesting that EH deviations over longer horizons are of limited relevance in economic terms. Out-of-sample, evidence against the EH is weaker because Θs are comparably small in absolute magnitudes and greater than zero only in 13 of 25 combinations.

Switching from the standard to the extended estimation also generates value to out-of-

[11] We repeated the empirical analysis using the performance fee of Fleming et al. (2001) and find qualitatively identical and quantitatively very similar results as we do for Θ (not reported to conserve space).

sample investors using the observable factor model (Panel B). While the in-sample results for standard and extended estimation are very similar again, the extended procedure outperforms the standard estimation with positive Θ values in 22 of 25 cases. Nevertheless, the model is not capable of beating the EH in economic terms, neither in-sample nor out-of-sample. Most premium returns relative to EH-consistent forecasts are negative.

Table 5 presents Θ estimates for power utility investors. The results are qualitatively identical to those for mean-variance investors, showing that our conclusions do not depend on assuming a specific utility function. The investor earns a premium return when she switches from the standard to the extended estimation procedure. Using the latent factor model, the investor can outperform the EH in-sample, but out-of-sample evidence is far less convincing. The observable factor model cannot beat the EH, neither in- or out-of-sample. Numerically, the results are somewhat more pronounced than those for quadratic utility, which suggests that $\sigma^* = 2\%$ *p.a.* in the mean-variance optimization leads to a more conservative allocation.

Overall, these results suggest that the information hidden to affine models estimated with the standard procedure but captured through the extended procedure results in economic gains for bond investors. For instance, out-of-sample, mean-variance investors with a one-year horizon would pay an annual premium of up to 3.5% to switch from the standard to the extended estimation of the latent factor model, and 1% when using the observable factor model. Power utility investors would be willing to pay even more. Relative to the EH, however, bond investors earn premium returns in-sample (when they use that latent factor model) but not out-of-sample.

4.4 Summary of results: Can ATSMs beat the EH?

Our results show that extending ATSM estimations beyond fitting yields to additionally match past excess returns captures information otherwise unspanned or hidden to standard ATSM

estimations. The extension leads to a substantial improvement in forecast accuracy for bond excess returns and to economic gains for bond portfolio investors.

We therefore employ this extended estimation to challenge the EH postulate of constant risk premia, where we use the historical average bond excess return as an EH-consistent benchmark predictor. While the models mostly outperform EH forecasts in terms of statistical accuracy, investors cannot systematically gain economic value from model forecasts out-of-sample. Our results reveal a contrast on the usefulness of ATSMs relative to the EH judged by statistical or economic criteria.

On the whole, our wealth of results can be catalyzed to the conclusion that ATSMs generally cannot beat the EH out-of-sample in terms of economic value. The finding that bond investors cannot systematically benefit from using conditional risk premia as compared to using the historical average can be viewed as the bond market analogue to the result of Goyal and Welch (2008) for stock markets.

5 Discussion of Results and Further Analysis

Viewed in isolation, our results may allow for different conclusions on the validity of the EH if one only considered in- or out-of-sample results or only statistical accuracy or economic value measures. As such, these - apparently - conflicting results, call for a deeper discussion.

5.1 Estimation following Adrian et al. (2013)

First, we repeat our empirical analysis using the regression-based ATSM estimation proposed by Adrian et al. (2013) to generate predictions for bond risk premia over a horizon of one month. Their approach also incorporates information about bond excess returns and, by construction,

fits realized bond excess returns almost perfectly. The model generates unbiased in-sample predictions of future excess returns for bonds with maturities from six months to ten years (Panel A in Table 6), with slope coefficients between 0.99 and 1.01 and regression-R^2 in the range from 4% to 6%. Panels B and C shows that the model has predictive accuracy in-sample, which translates into economic value in half of the cases. Out-of-sample the model's forecast accuracy is limited and economic value relative to the EH is negative. These results as such, as well as a comparison with the standard and extended estimation results presented above, provide further evidence for the tension between predictive accuracy versus economic value and in-sample (over-)fitting versus out-of-sample performance.

5.2 Statistical accuracy versus economic value

While many papers on the predictability of bond risk premia are concerned with statistical forecast accuracy, statistical accuracy *per se* does not imply economic value for bond investors. Our results indeed suggest conflicting conclusions about the validity of the EH based on statistical and economic criteria. For instance, using the observable factor model, we would reject the EH based on metrics of forecast accuracy but the same forecasts lead to economic losses compared to EH-implied constant risk premia. Similarly, we find for our latent factor model and for the Adrian et al. (2013) model various cases where the model beats the EH statistically but not economically and vice versa. Below we present general, model-free arguments as to why there may be a contrast between statistical and economic significance and evaluate our model results along these lines. These arguments are also useful when interpreting results of other papers that study the predictability of bond risk premia using various forecasting approaches.

5.2.1 Economic Relevance of EH Deviations

One reason for apparently conflicting results is that departures from the EH might be statistically significant but too small to be exploited by bond investors. Since there is no "natural" upper bound for economic value measures (similar to a regression R^2 capped by one or forecast errors floored by zero), we compare the economic performance of model forecasts to the performance of the same strategy under perfect foresight. If perfect foresight returns of the strategy are high but the model evaluated only captures a (small) fraction of these excess returns, EH deviations are not exploited because the model fails. If the model captures a large fraction of perfect foresight returns but returns are nevertheless economically small, this suggests that "true" EH deviations are indeed economically irrelevant.[12, 13]

To get a feeling for the economic relevance of EH deviations, we plot average excess returns of buy-and-hold investors and perfect foresight portfolios in Figure 1. Buy-and-hold excess returns capturing constant risk premia increase with maturity and decrease with forecast horizon. The patterns are very similar for perfect foresight investors but with average excess returns on a higher level. It is more valuable for investors to accurately predict short-horizon as compared to long-horizon bond excess returns. For instance, investors that buy and hold the long-term bond ($T-\tau = 60$ months) over horizons τ =1, 12, and 60 months earn average excess returns of 1.86%, 1.31%, and 0.59% *p.a.*. The perfect foresight excess returns for the same combinations are 10.82%, 3.46%, and 1.65% *p.a.* This shows that EH deviations are less important for

[12]For a simple strategy that just goes long (short) when the expected excess return is positive (negative), the returns based on model forecasts relative to perfect foresight are bounded by plus/minus one. For optimal portfolios, model-based returns could exceed those of perfect-foresight portfolios, which would imply a less than optimal risk-return trade-off. Repeating this empirical exercise with model and perfect foresight Θs leads to qualitatively the same conclusions that we report for returns below.

[13]Even models that perfectly capture risk premia may not generate an economic performance equal to that based on perfect foresight because departures from the EH may not be exclusively driven by (predictable) risk premia. Similarly, in the presence of noise or other determinants of EH deviations, it would not be possible to achieve an R^2 of 1 with perfect risk premium predictions in regressions of realized excess returns.

increasing τ and that having a less then perfect forecast model for short horizons may add more economic value than a perfect forecast model for longer horizons.

In Figure 2, we plot the excess returns of portfolios allocated using forecasts based on constant risk premia (in light gray), the standard estimation (in dark gray), and the extended estimation (in black) relative to perfect foresight portfolio returns. The graphs show that EH deviations are not as important economically as statistical results might suggest because constant risk premium forecasts capture a large fraction of perfect foresight returns. The fraction captured by extended estimation forecasts generally exhibits similar patterns as regression R^2s and $R2$-statistics in Tables 2 and 3; for instance, we typically see the highest R^2s, $R2$s, and fractions of perfect foresight returns captured at the 12-month horizon. In contrast, and particularly pronounced in-sample, the economic value decreases with horizon (Table 4), consistent with comparably lower statistical accuracy at short horizons adding higher economic value than more accurate forecasts for longer horizons. In other words, statistical accuracy cannot lead to economic value when EH deviations are too small to be exploited by investors.

As a note of caution, we emphasize the illustrative nature of the exercise carried out in this subsection. It may well be that more complex settings lead to qualitatively different conclusions, but our example serves to show one case where the discrepancy between statistical and economic value metrics is easy to rationalize.

5.2.2 Information in Economic Value versus Statistical Accuracy Measures

Conflicting conclusions based on metrics of statistical accuracy and economic value may also result from the construction of the measures used. Common measures of predictive ability are based on loss functions involving squared or absolute forecast errors which, by definition, ignore the sign of forecast errors. Getting the sign right, however, is of utmost importance for

investors since it determines whether to take a long versus a short position or whether to invest in the risky asset versus the risk-free asset.

As a measure of directional accuracy, we compute hit ratios measuring the fraction of correctly signed forecasts. Table 7 reports the hit ratios of the extended estimation relative to the hit ratios of constant risk premium forecasts, with asterisks (circles) indicating that model hit ratios are significantly higher (lower) than those of constant risk premium forecasts. The results confirm that our finding that the economic value analysis is more in favor of the EH than the statistical accuracy results can partly be explained by forecasts having small squared/absolute errors but nonetheless pointing in the wrong direction. This can best be seen for the observable factor model, which beats the EH in terms of forecast accuracy but nonetheless does not generate economic value (Tables 3 and 4). This is consistent with the model getting the direction right only in 9 (5) of 25 cases in-sample (out-of-sample).

To further gauge the relation between statistical versus economic significance, we plot constant risk premium forecast errors (black circles) and model forecast errors (red crosses) against realized excess returns in Figure 3. The shaded areas represent scenarios where forecasts have the wrong sign and hence forecast errors that lead to bond portfolio losses. While standard predictive ability measures only account for the distribution of forecast errors across the x-axis in absolute terms, the economic value for investors depends on the signed forecast errors' joint distribution with realizations (on the y-axis). The distribution across the x-axis looks relatively similar for model and constant risk premium forecasts but the model forecast errors exhibit a larger dispersion across the y-axis. These patterns explain why statistical predictability does not (necessarily) map into economic gains for bond investors.

5.3 In-Sample versus Out-of-Sample Results

Our finding that the extended estimation dominates the standard estimation is robust in- and out-of-sample, for both the latent and the observable factor model. The observable factor model suggests identical conclusions on the statistical and economic (ir-)relevance of EH deviations in- and out-of-sample. For the latent factor model, however, the extended estimation delivers very strong statistical and economic results against the EH in-sample, which do not survive the out-of-sample test.

As Duffee (2010, page 1) states, "Flexibility and overfitting go hand-in hand" when evaluating ATSMs in-sample, and thus studying out-of-sample properties is warranted to gauge the extent to which bond excess returns are predictable. The prevalent parameter uncertainty in ATSMs (e.g. Feldhütter et al., 2012) is a potential source of overfitting and the latent factor model's additional flexibility of state variables being estimated (rather than observed) apparently only helps in-sample. The challenge for future research is to consider modeling approaches that are flexible but limit overfitting. One conceivable route may be to impose economically reasonable restrictions, for instance, by using the EH as an economic anchor or as a prior in estimation.

5.4 Macroeconomic Uncertainty

Building on our discussion above, we now explore whether the model's out-of-sample performance depends on how the economic relevance of EH deviations varies over time with macro uncertainty. We find that extended estimation forecasts of the latent factor model beat the EH out-of-sample when there is high uncertainty about the future state of the economy.

We measure uncertainty using the data of Jurado et al. (2015) and classify our sample into

periods of very high, high, low, and very low uncertainty using the 90%, 75%, 25%, and 10% quantiles of their one-month and twelve-month uncertainty measures. Our results in Table 8 suggest that the extended estimation typically generates large economic gains relative to the standard estimation, only when short-term uncertainty is very low results become less clear. Compared to the EH, we find that the model beats the EH when forecasting excess returns of bonds with a maturity of one year or longer during periods of very high uncertainty. All measures of economic value are positive, increase with bond maturity, and decrease with forecast horizon. Qualitatively similar but quantitatively less pronounced patterns also apply during periods of high uncertainty. By contrast, we find that the superiority of model forecasts relative to the EH deteriorates when uncertainty is low and that the model's economic value is mostly negative when short-term uncertainty is very low.

These results are consistent with the notion that deviations from the EH are more likely to occur in periods of heightened volatility. In other words, the time-variation in the risk premium captured by our model is related to the degree of macroeconomic uncertainty. During periods of (very) low uncertainty, the model appears to add relatively more noise than economically relevant return information beyond EH forecasts. By contrast, the model conveys valuable information for future bond excess returns in periods of high uncertainty about the macroeconomy. Specifically, our out-of-sample results suggest that investors would be willing to pay a sizeable premium to switch from EH- to model-forecasts when uncertainty is high.[14]

Our findings are, thus, consistent with previous research showing that macro factors are informative for bond risk premia (e.g., Ludvigson and Ng, 2009) and suggest that more research is

[14]We focus here on the latent factor model because our results above suggest that it is more successful in generating economic; see Tables 4 and 5. For the model with observable factors, we also find that the economic value of the extended estimation forecasts compared to the standard estimation forecasts and relative to the EH appears related to macro uncertainty. This evidence, however, is more mixed, which suggests that the additional flexibility from modeling state variables is helpful for capturing uncertainty-related EH deviations when generating out-of-sample forecasts.

warranted on how the economic value of bond risk premium forecasts are related to uncertainty about the economy. Recently, Gargano et al. (2015) also provide evidence that economic gains depend on the state of the economy. Going forward, it seems natural to consider models of the term structure which switch from a simple EH anchor in calm times to ATSM specifications in more turbolent times, with uncertainty acting as the state variable that drives the switch from one extreme to another.

5.5 Additional Results and Robustness Checks

To corroborate our findings, we perform various robustness checks and additional empirical analyses. In the Internet Appendix, we summarize evidence on alternative ATSM specifications and discuss our results in relation to forward rates-based predictions of bond excess returns. Furthermore, we show that our conclusions are robust through the recent financial crisis and that they apply uniformly to Japan, Switzerland, Germany, and the UK as well.

6 Conclusion

In this paper, we offer new insights on the expectations hypothesis (EH) by studying the economic benefits that accrue to bond portfolio investors who exploit predictable deviations from the EH. We estimate conditional bond risk premia using affine term structure models (ATSMs) by employing a novel estimation strategy that jointly fits the term structure of model yields to the observed yield curve and additionally matches model risk premia with bond excess returns observed in the past. This extended procedure allows investors to capture predictive information beyond the cross section of yields (i.e. unspanned by the term structure) and to update beliefs about the model's predictive ability based on its past performance. To evaluate

the model against the EH, we use averages of historical bond excess returns to consistently estimate constant risk premia as postulated by the EH.

We find that, for 25 combinations of horizons and maturities ranging from one month to ten years, the extended estimation captures predictive information otherwise hidden to standard ATSM estimations. However, while portfolios based on model-forecasts earn positive excess returns, they perform worse than corresponding EH benchmark portfolios in out-of-sample analysis. The apparent wedge in conclusions from statistical and economic assessments of the EH is not rooted in the use of ATSMs but, as we show, potentially applies to other approaches for predicting bond risk premia. The bottom line is that even models with high regression R^2s or predictive ability cannot guarantee to provide bond investors with economic gains relative to presuming that the EH holds.

Overall, our results suggest that the EH presumption of constant risk premia, while being statistically rejected by the data, still provides a useful approximation for the out-of-sample behavior of bond excess returns, especially for the purpose of fixed income asset allocation over longer forecast horizons and during times of low uncertainty.

References

Adrian, T., Crump, R. K., and Moench, E. (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics*, 110(1):110–138.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.

Backus, D. K., Foresi, S., Mozumdar, A., and Wu, L. (2001). Predictable changes in yields and forward rates. *Journal of Financial Economics*, 59:281–311.

Balduzzi, P. and Chang, I.-H. E. (2012). A simple test of the affine class of term structure models. *Review of Asset Pricing Studies*, 2:203–244.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance*, 55:225–264.

Barillas, F. (2013). Can we exploit predictability in bond markets? Working paper, Emory University.

Bekaert, G. and Hodrick, R. J. (2001). Expectations hypotheses tests. *Journal of Finance*, 56:1357–1394.

Campbell, J. Y. and Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies*, 58:495–514.

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

Cheridito, P., Filipović, D., and Kimmel, R. (2007). Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics*, 83:123–170.

Cieslak, A. and Povala, P. (2013). Expected returns in treasury bonds. Working Paper, Northwestern University and University of Lugano.

Cochrane, J. and Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95:138–160.

Collin-Dufresne, P., Goldstein, R. S., and Jones, C. S. (2008). Identification of maximal affine term structure models. *Journal of Finance*, 63:743–795.

Dai, Q. and Singleton, K. J. (2000). Specification analysis of affine term structure models. *Journal of Finance*, 55:1943–1978.

Dai, Q. and Singleton, K. J. (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics*, 63:415–441.

Della Corte, P., Sarno, L., and Thornton, D. L. (2008). The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value. *Journal of Financial Economics*, 89:158–174.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13:253–263.

Duffee, G. (2010). Sharpe ratios in term structure models. Working paper, Johns Hopkins University.

Duffee, G. (2011). Information in (and not in) the term structure. *Review of Financial Studies*, 24:2895–2934.

Duffie, D., Filipović, D., and Schachermayer, W. (2003). Affine processes and applications in finance. *Annals of Applied Probability*, 13:984–1053.

Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6:379–406.

Fama, E. F. (1984). The information in the term structure. *Journal of Financial Economics*, 13:509–528.

Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates. *American Economic Review*, 77:680–692.

Feldhütter, P. and Lando, D. (2008). Decomposing swap spreads. *Journal of Financial Economics*, 88:375–405.

Feldhütter, P., Larsen, L., Munk, C., and Trolle, A. (2012). Keep it simple: Dynamic bond portfolios under parameter uncertainty. Working Paper, London Business School.

Filipović, D., Mayerhofer, E., and Schneider, P. (2013). Density approximations for multivariate affine jump-diffusion processes. *Journal of Econometrics*. forthcoming.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *Journal of Finance*, 56:329–352.

Gargano, A., Pettenuzzo, D., and Timmermann, A. G. (2015). Bond return predictability: Economic value and links to the macroeconomy. *Working Paper*.

Goetzmann, W., Ingersoll, J., Spiegel, M., and Welch, I. (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies*, 20:1503–1546.

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

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The U.S. treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54:2291–2304.

Hall, P., Horowitz, J., and Jing, B. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, 82:561–574.

Hansen, L. P. and Hodrick, R. J. (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *The Journal of Political Economy*, pages 829–853.

Joslin, S., Priebsch, M., and Singleton, K. (2013). Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance*. fortchoming.

Joslin, S., Singleton, K., and Zhu, H. (2011). A new perspective on gaussian dynamic term structure models. *Review of Financial Studies*, 24:926–970.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *The American Economic Review*, 105(3):1177–1216.

Leitch, G. and Tanner, E. (1991). Economic forecast evaluation: Profit versus the conventional error measures. *American Economic Review*, 81:580–590.

Litterman, R. and Scheinkman, J. A. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, 1:54–61.

Ludvigson, S. and Ng, S. (2009). Macro factors in bond risk premia. *Review of Financial Studies*, 22:5027–5067.

Mueller, P., Vedolin, A., and Zhou, H. (2011). Short run bond risk premia. Discussion Paper 686, Financial Markets Group, LSE.

Newey, W. and West, K. (1987). A simple positive semi-definite, heteroscedasticity and autorcorrelation consitent covariance matrix. *Econometrica*, 55:703–708.

Patton, A., Politis, D., and White, H. (2009). Correction to "automatic block-length selection for dependent bootstrap". *Econometric Reviews*, 28:372–375.

Politis, D. and White, H. (2004). Automatic block-length selection for dependent bootstrap. *Econometric Reviews*, 23:53–70.

Sarno, L., Thornton, D. L., and Valente, G. (2007). The empirical failure of the expectations hypothesis of the term structure of bond yields. *Journal of Financial and Quantitative Analysis*, 42:81–100.

Thornton, D. and Valente, G. (2012). Out-of-sample predictions of bond excess returns and forward rates: an asset allocation perspective. *Review of Financial Studies*, 25:3141–3168.

West, K., Edison, H., and Cho, D. (1993). A utility based comparison of exchange rate volatility. *Journal of International Economics*, 35:23–45.

Table 1: Yield Pricing Errors

Panel A: $A_0(3)$ **Model with Latent Factors**

Yield Maturity		All T	1m $\leq T <$ 12m	12m $\leq T <$ 60m	60m $\leq T \leq$ 120m
Number of Yields		24.00	7.00	10.00	7.00
RMSE	standard estimation	16.94	30.19	5.57	5.29
	extended estimation	26.13	45.57	10.80	9.86
Sd	standard estimation	16.86	29.26	5.44	5.29
	extended estimation	26.12	45.37	10.78	9.67

Panel B: $A_0(3)$ **Model with Observable Factors**

Yield Maturity		All T	1m $\leq T <$ 12m	12m $\leq T <$ 60m	60m $\leq T \leq$ 120m
Number of Yields		24.00	7.00	10.00	7.00
RMSE	standard estimation	8.23	11.16	6.42	6.98
	extended estimation	8.22	11.15	6.42	6.99
Sd	standard estimation	8.23	11.12	6.32	6.95
	extended estimation	8.23	11.11	6.32	6.96

Notes: The Table summarizes root mean squared yield pricing errors (RMSEs) and standard deviations of yield pricing errors for the standard estimation (the estimation procedure only fitting yields) and the extended estimation (the estimation procedure fitting yields and matching model risk premia to bond excess returns observed in the past) of the $A_0(3)$ model. Panel A contains results for the latent factor model, Panel B for the observable factor model. We estimate the models using monthly US data from January 1952 to December 2012.

Table 2: Time-Varying Risk Premium Regressions

Panel A: $A_0(3)$ Model with Latent Factors

Standard Estimation Procedure

		$T-\tau$	1m	3m	12m	24m	60m
τ	1m	b	0.50	0.82	1.02	1.08*	1.05**
		se	(0.50)	(0.57)	(0.59)	(0.55)	(0.38)
		R^2	0.00	0.01	0.01	0.02	0.02
	3m	b	0.80	0.62	1.09	1.12*	1.02**
		se	(0.58)	(0.61)	(0.58)	(0.52)	(0.37)
		R^2	0.02	0.01	0.03	0.04	0.05
	12m	b	1.17**	1.28**	1.41**	1.32**	1.20**
		se	(0.41)	(0.42)	(0.43)	(0.40)	(0.32)
		R^2	0.10	0.11	0.14	0.15	0.19
	24m	b	0.85	0.89	0.91*	0.92*	0.96**
		se	(0.50)	(0.49)	(0.44)	(0.38)	(0.27)
		R^2	0.06	0.07	0.08	0.10	0.17
	60m	b	0.86*	0.86*	0.87*	0.85*	0.84**
		se	(0.39)	(0.38)	(0.37)	(0.34)	(0.29)
		R^2	0.09	0.10	0.10	0.11	0.13

Extended Estimation Procedure

		$T-\tau$	1m	3m	12m	24m	60m
τ	1m	b	0.72**	0.79**	0.76**	0.76**	0.81**
		se	(0.13)	(0.14)	(0.11)	(0.10)	(0.11)
		R^2	0.13	0.12	0.12	0.12	0.11
	3m	b	0.89**	0.88**	0.87**	0.86**	0.89**
		se	(0.15)	(0.14)	(0.10)	(0.09)	(0.09)
		R^2	0.29	0.28	0.32	0.32	0.30
	12m	b	1.03**	1.03**	1.06**	1.06**	1.11**
		se	(0.05)	(0.05)	(0.04)	(0.04)	(0.06)
		R^2	0.79	0.83	0.87	0.87	0.80
	24m	b	1.05**	1.06**	1.08**	1.08**	1.12**
		se	(0.10)	(0.09)	(0.09)	(0.10)	(0.13)
		R^2	0.69	0.70	0.72	0.73	0.71
	60m	b	0.76**	0.73**	0.78**	0.83**	0.91**
		se	(0.13)	(0.13)	(0.13)	(0.13)	(0.12)
		R^2	0.23	0.22	0.24	0.27	0.32

Panel B: $A_0(3)$ Model with Observable Factors

Standard Estimation Procedure

		$T-\tau$	1m	3m	12m	24m	60m
τ	1m	b	0.90**	0.83**	0.92**	1.09**	0.92**
		se	(0.19)	(0.20)	(0.23)	(0.32)	(0.31)
		R^2	0.06	0.05	0.03	0.03	0.02
	3m	b	1.03**	0.94**	1.09**	1.22**	0.91**
		se	(0.24)	(0.26)	(0.30)	(0.37)	(0.31)
		R^2	0.15	0.10	0.08	0.07	0.05
	12m	b	1.03**	1.03**	1.18**	1.18**	1.05**
		se	(0.36)	(0.39)	(0.42)	(0.39)	(0.30)
		R^2	0.15	0.12	0.12	0.13	0.18
	24m	b	0.99**	0.96*	0.92*	0.89**	0.89**
		se	(0.37)	(0.38)	(0.38)	(0.33)	(0.23)
		R^2	0.12	0.10	0.09	0.10	0.17
	60m	b	0.85**	0.86**	0.87**	0.83**	0.81**
		se	(0.29)	(0.30)	(0.31)	(0.30)	(0.28)
		R^2	0.15	0.15	0.14	0.14	0.16

Extended Estimation Procedure

		$T-\tau$	1m	3m	12m	24m	60m
τ	1m	b	0.85**	0.79**	0.93**	1.11**	0.91**
		se	(0.18)	(0.19)	(0.23)	(0.33)	(0.31)
		R^2	0.07	0.05	0.03	0.03	0.02
	3m	b	0.94**	0.88**	1.08**	1.22**	0.91**
		se	(0.22)	(0.25)	(0.31)	(0.37)	(0.31)
		R^2	0.15	0.10	0.08	0.07	0.05
	12m	b	0.92**	0.93*	1.14**	1.16**	1.04**
		se	(0.33)	(0.37)	(0.41)	(0.39)	(0.29)
		R^2	0.15	0.12	0.12	0.13	0.18
	24m	b	0.92**	0.91**	0.91*	0.87**	0.88**
		se	(0.33)	(0.34)	(0.36)	(0.32)	(0.24)
		R^2	0.13	0.11	0.09	0.10	0.17
	60m	b	0.81**	0.83**	0.85**	0.82**	0.81**
		se	(0.27)	(0.28)	(0.29)	(0.30)	(0.28)
		R^2	0.15	0.15	0.15	0.14	0.16

Notes: The Table presents results for regressing realized bond excess returns on risk premia implied by the $A_0(3)$ model using the standard estimation procedure that only fits yields and using the extended estimation procedure that fits yields and matches model risk premia to bond excess returns observed in the past. Panel A contains results for the latent factor model, Panel B for the observable factor model. We report estimates for 25 horizon (τ) and maturity ($T-\tau$) combinations. The horizons are indicated in the rows, the maturities in the columns. b is the estimate of the slope coefficient. ** and * indicates that the estimate differs from zero at the 99% or 95% level, respectively. Significance is assessed using standard errors (se) based on Hansen and Hodrick (1980). R^2 denotes the in-sample coefficient of determination. We estimate the models using monthly US data from January 1952 to December 2012.

Table 3: Statistical Accuracy of Bond Excess Return Forecasts

Panel A: $A_0(3)$ Model with Latent Factors

	Full Sample										Out-of-Sample									
	Extended vs. Standard Estimation					Model vs. EH Forecast					Extended vs. Standard Estimation					Model vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	0.21*	0.14*	0.09*	0.09*	0.09*	0.11*	0.11*	0.1*	0.11*	0.11*	−0.27	−0.23	−0.26	−0.25°	−0.16	−0.33°	−0.27°	−0.27°	−0.26°	−0.18°
3m	0.33*	0.29*	0.28*	0.28*	0.26*	0.28*	0.27*	0.31*	0.31*	0.3*	−0.14	−0.12	−0.13	−0.12	−0.06	−0.26	−0.22	−0.23	−0.23	−0.16
12m	0.77*	0.81*	0.85*	0.85*	0.74*	0.79*	0.83*	0.86*	0.87*	0.79*	0.26*	0.33*	0.37*	0.39*	0.36*	0.08	0.17	0.22	0.23*	0.21*
24m	0.67*	0.68*	0.69*	0.7*	0.64*	0.69*	0.7*	0.71*	0.73*	0.7*	0.22	0.27	0.35*	0.4*	0.39*	0.07	0.13	0.21	0.25*	0.23*
60m	0.13	0.11	0.14	0.18	0.21*	0.21	0.19	0.22*	0.26*	0.31*	0.04	0.06	0.17	0.24*	0.28*	−0.2	−0.18	−0.05	0.02	0.05

Panel B: $A_0(3)$ Model with Observable Factors

	Full Sample										Out-of-Sample									
	Extended vs. Standard Estimation					Model vs. EH Forecast					Extended vs. Standard Estimation					Model vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	0.01*	0	0	0	0	0.02	0.04	0.03	0.03*	0.02*	0.03	0.03*	0.04*	0.03*	0.02*	0.01	0.03	0.04	0.03*	0.01
3m	0.01	0	0	0	0	0.14*	0.1	0.08*	0.07*	0.05*	0.07	0.08*	0.08*	0.07*	0.08*	0.12	0.1	0.09*	0.08*	0.05
12m	0	−0.01	0	0	0	0.15	0.12	0.12*	0.13*	0.18*	0.19*	0.19*	0.18*	0.18*	0.23*	0.18	0.17*	0.17*	0.19*	0.25*
24m	0.01	0.01	0	0	0	0.13	0.11	0.09	0.1	0.17	0.21*	0.21*	0.21*	0.22*	0.29*	0.14	0.13	0.12	0.13	0.22
60m	0	0	0	0	0	0.14	0.15	0.14	0.13	0.15	0.29*	0.29*	0.28*	0.3*	0.34*	0.04	0.04	0.02	0.03	0.08

Notes: The Table presents $R2$ statistics measuring the predictive ability of the ATSM estimated with the extended procedure relative to the standard estimation and EH-consistent constant risk premium benchmarks. The $R2$ statistic is defined as $R2 \equiv 1 - MSE^m/MSE^b$ where MSE^m and MSE^b denote the mean squared errors of the model and benchmark forecasts, respectively. We report in-sample and out-of-sample results for 25 horizon (τ) and maturity ($T-\tau$) combinations for the latent and observable factor model in Panels A and B, respectively. The horizons are indicated in the rows, the maturities in the columns. * ($^\circ$) identifies $R2$s that are significantly positive (negative) as judged by confidence intervals obtained as the 5%- and 95%-percentiles from a block bootstrap procedure. For the in-sample analysis, we estimate the models using monthly US data from January 1952 to December 2012. In the out-of-sample analysis, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t.

Table 4: Economic Value of Bond Excess Return Forecasts for Mean-Variance Investors

Panel A: $A_0(3)$ Model with Latent Factors

	Full Sample															Out-of-Sample													
	Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast					Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	54	107	167	255	386	5	11	64	120	186	-5	-7	43	120	223	39	82	131	164	276	-8	2	70	90	182	-19	-30	10	19	-45
3m	36	84	164	243	371	2	13	76	136	200	-2	1	75	135	230	30	70	124	163	286	-2	14	133	187	382	-8	-9	38	36	81
12m	21	59	143	212	314	4	17	70	125	142	4	20	76	124	192	17	48	99	148	241	6	25	113	195	353	0	7	31	50	60
24m	14	40	95	139	213	3	10	36	64	76	3	11	39	61	102	12	31	63	101	195	2	8	43	107	273	0	1	5	14	38
60m	6	16	45	64	93	0	1	6	14	21	0	1	7	12	21	7	18	46	53	60	1	5	31	65	158	-0	-1	-4	-15	-39

Panel B: $A_0(3)$ Model with Observable Factors

	Full Sample															Out-of-Sample													
	Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast					Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	55	103	108	120	183	0	0	2	2	21	-4	-11	-13	-14	21	50	86	64	159	241	1	4	-14	66	87	-9	-25	-57	15	-70
3m	36	75	82	119	163	-0	-1	-6	5	3	-2	-8	-3	14	21	33	63	99	120	199	-2	-0	31	42	82	-6	-16	14	-4	-5
12m	16	35	50	81	148	-0	-0	-0	0	3	-1	-2	-10	0	34	16	39	73	115	243	2	6	28	47	107	-1	-1	6	19	51
24m	11	26	36	59	117	-0	-0	1	-1	-2	-0	-2	-13	-11	14	12	28	44	63	127	2	5	12	24	72	1	-2	-13	-23	-40
60m	6	15	32	38	61	-0	0	-1	-2	0	-0	-0	-3	-8	-2	6	15	27	23	21	1	3	13	27	74	-1	-3	-21	-48	-128

Notes: The Table presents results on the economic value generated by ATSM forecasts using the extended estimation procedure. We report results of the in-sample and out-of-sample analysis for 25 horizon (τ) and maturity ($T-\tau$) combinations using the latent and observable factor model in Panels A and B, respectively. The horizons are indicated in the rows, the maturities in the columns. In each Panel, we first present the excess returns of optimal bond portfolios using extended estimation forecasts. For an investment horizon τ, an investor with target volatility of $\sigma^* = 2\%$ and relative risk aversion $\rho = 3$ optimally allocates his wealth between the risk-free bond with maturity τ and the risky bond with maturity $T > \tau$. In the remaining columns, we report the economic value added by the extended compared to the standard estimation and compared to the EH-consistent constant risk premium forecasts using the performance measure of Goetzmann et al. (2007). All performance measures are annualized and reported in basis points. For the in-sample analysis, we estimate the models using monthly US data from January 1952 to December 2012. In the out-of-sample analysis, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t.

Table 5: Economic Value of Bond Excess Return Forecasts for Power Utility Investors

Panel A: $A_0(3)$ Model with Latent Factors

	Full Sample															Out-of-Sample													
	Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast					Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	54	106	258	467	906	5	10	79	186	345	-5	-9	59	185	440	39	82	205	313	564	-8	0	72	89	138	-19	-33	-6	-21	-140
3m	36	84	264	466	882	2	11	109	220	344	-2	0	105	223	460	30	75	201	322	579	-2	17	166	282	508	-8	-10	30	39	23
12m	21	59	215	384	739	4	18	99	190	280	4	20	104	198	390	17	46	153	267	512	6	25	138	264	527	0	5	35	58	65
24m	14	40	141	251	492	3	10	51	98	148	3	11	51	93	192	11	31	99	187	400	2	9	57	141	395	-0	1	7	15	38
60m	6	16	57	102	200	0	1	8	19	29	-0	1	7	15	23	6	17	56	92	126	1	5	34	76	199	-0	-1	-4	-16	-75

Panel B: $A_0(3)$ Model with Observable Factors

	Full Sample															Out-of-Sample													
	Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast					Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	55	103	181	244	448	0	0	-2	-2	11	-4	-11	-15	-36	-18	50	90	132	317	579	1	5	-1	92	146	-8	-25	-79	-17	-106
3m	36	74	154	259	425	-0	-1	-1	5	1	-2	-9	0	21	-6	33	68	174	262	493	-1	-0	47	55	116	-5	-17	5	-17	-87
12m	16	35	89	172	383	-0	-1	-2	0	-1	-1	-3	-13	-3	44	16	38	124	231	525	1	5	32	61	143	-1	-2	7	19	21
24m	10	25	64	127	291	-0	-0	-1	0	-4	-0	-2	-16	-17	0	12	28	79	147	319	2	5	14	34	89	0	-2	-13	-27	-75
60m	5	14	42	70	141	-0	-0	0	0	0	-0	-1	-4	-9	-16	6	14	40	62	114	1	3	15	30	83	-1	-3	-20	-51	-177

Notes: The Table presents results on the economic value generated by ATSM forecasts using the extended estimation procedure. We report results for the in-sample and out-of-sample analysis in Panels A and B for 25 horizon (τ) and maturity ($T-\tau$) combinations, respectively. The horizons are indicated in the rows, the maturities in the columns. In each Panel, we first present the excess returns of optimal bond portfolios using extended estimation forecasts. For an investment horizon τ, an investor with target volatility of $\sigma^* = 2\%$ and relative risk aversion $\rho = 3$ optimally allocates his wealth between the risk-free bond with maturity τ and the risky bond with maturity $T > \tau$. In the remaining columns, we report the economic value added by the extended compared to the standard estimation and compared to the EH-consistent constant risk premium forecasts using the performance measure of Goetzmann et al. (2007). All performance measures are annualized and reported in basis points. For the in-sample analysis, we estimate the models using monthly data from September 1989 for Japan, January 1988 for Switzerland, and April 1987 for Germany, UK, and US. The sample period ends in March 2011 for all countries. For the US we also report results for a longer data set covering the period from January 1952 to December 2003. In the out-of-sample analysis, we estimate for each country data set the first model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t.

Table 6: ATSM Estimation following Adrian, Crump, and Moench (2013)

Panel A: Time-Varying Risk Premium Regressions

	6m	12m	18m	24m	30m	36m	42m	48m	54m	60m	84m	120m
b	0.99**	1.01**	1.00**	0.99**	1.00**	1.00**	1.00**	1.00**	1.00**	1.00**	1.00**	1.00**
se	(0.32)	(0.28)	(0.31)	(0.29)	(0.31)	(0.28)	(0.25)	(0.22)	(0.21)	(0.20)	(0.19)	(0.22)
R^2	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.06	0.05

Panel B: Statistical Accuracy of Bond Excess Return Forecasts compared to EH Forecasts

	6m	12m	18m	24m	30m	36m	42m	48m	54m	60m	84m	120m
Full Sample	0.05	0.05	0.04	0.04	0.04⋆	0.04⋆	0.04⋆	0.04⋆	0.05⋆	0.05⋆	0.06⋆	0.05⋆
Out-of-Sample	−0.02	−0.01	−0.02	−0.03	−0.02	−0.02	−0.01	−0.01	0	0.01	0.02	0.01

Panel C: Economic Value of Bond Excess Return Forecasts for Mean-Variance Investors

	6m	12m	18m	24m	30m	36m	42m	48m	54m	60m	84m	120m
Full Sample												
Excess Returns	141	113	86	80	71	75	79	83	86	99	99	95
Model vs. EH	−14	−7	−17	−13	−17	−7	1	9	16	30	40	47
Out-of-Sample												
Excess Returns	110	63	37	42	57	51	54	52	33	56	99	51
Model vs. EH	−45	−57	−69	−58	−43	−56	−65	−59	−64	−77	−10	−41

Notes: We use the estimation procedure proposed by Adrian et al. (2013) to forecast of one-month excess returns of bonds with maturities between six months and ten years. For the in-sample analysis, we estimate the models using monthly US data from January 1952 to December 2012. In the out-of-sample analysis, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t. Panel A presents results for in-sample regressions of realized bond excess returns on model-implied risk premia b is the estimate of the slope coefficient. ** and * indicates that the estimate differs from zero at the 99% or 95% level, respectively. Significance is assessed using standard errors (se) based on Hansen and Hodrick (1980). R^2 denotes the in-sample coefficient of determination. Panel B presents $R2$ statistics measuring the predictive ability of the model relative to the EH-consistent constant risk premium benchmarks. The $R2$ statistic is defined as $R2 \equiv 1 - MSE^m/MSE^b$ where MSE^m and MSE^b denote the mean squared errors of the model and benchmark forecasts, respectively. We report in-sample and out-of-sample results and ⋆ identifies $R2$s that are significantly positive as judged by confidence intervals obtained as the 5%- and 95%-percentiles from a block bootstrap procedure. Panel C presents results on the economic value generated by the model relative to the EH-consistent constant risk premium forecast. An investor with target volatility of $\sigma^* = 2\%$ and relative risk aversion $\rho = 3$ optimally allocates his wealth between the risk-free bond with one-month maturity and the risky bond with maturity indicated in the column header. We measure economic value by the performance measure of Goetzmann et al. (2007). All performance measures are annualized and reported in basis points.

Table 7: Directional Accuracy of Bond Excess Return Forecasts

Panel A: $A_0(3)$ Model with Latent Factors

	Full Sample					Out-of-Sample				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	0.9°	0.89°	1.01	$1.08^{\star}$	$1.13^{\star\star}$	$0.87^{\circ\circ}$	$0.87^{\circ\circ}$	0.96	0.95	0.95
3m	0.89°	0.92°	$1.18^{\star\star}$	$1.22^{\star\star}$	$1.31^{\star\star}$	$0.88^{\circ\circ}$	0.9°	1.07	1.03	1.05
12m	$1.2^{\star\star}$	$1.3^{\star\star}$	$1.49^{\star\star}$	$1.54^{\star\star}$	$1.52^{\star\star}$	1	1.05	1.09	$1.12^{\star}$	1.09
24m	$1.18^{\star}$	$1.27^{\star\star}$	$1.31^{\star\star}$	$1.35^{\star\star}$	$1.35^{\star\star}$	0.98	0.97	1.03	1.09	1.14
60m	1.03	1.08	1.14	$1.27^{\star}$	$1.44^{\star}$	0.99	1.01	1.01	0.97	0.94

Panel B: $A_0(3)$ Model with Observable Factors

	Full Sample					Out-of-Sample				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	0.91°	0.9°	0.93°	0.98	1.06	$0.83^{\circ\circ}$	$0.81^{\circ\circ}$	$0.86^{\circ\circ}$	0.99	0.95
3m	0.91°	0.87°	0.94	1.02	1.08	$0.83^{\circ\circ}$	$0.8^{\circ\circ}$	0.96	1	1.01
12m	0.94	0.94	0.96	1.05	1.12	0.93	0.91	1	1.03	1.07
24m	0.98	0.96	0.97	1.03	1.11	1.02	0.96	1.02	0.99	1
60m	0.97	0.97	0.96	1.05	$1.31^{\star}$	0.96	0.96	0.93	0.89	0.91

Notes: The Table presents measures of the directional accuracy of extended estimation compared to EH-consistent constant risk premium forecasts based on hit ratios that measuring the ratio of correctly signed forecasts. Panels A and B contain results for the latent and observable factor model, respectively. Values reported are computed as the fraction of the ATSM model hit ratio relative to the EH hit ratio. Estimates with a * indicate that the model hit ratio exceeds the 95%-percentile of the bootstrapped constant risk premium hit ratio distribution. ** indicates that the 5%-percentile of the model distribution is higher than the 95%-percentile of the constant risk premium hit ratio distribution. $^{\circ}$ and $^{\circ\circ}$ indicate analogous results when the model hit ratio is lower than the constant risk premium hit ratio. We present in- and out-of-sample results for 25 horizon (τ) and maturity ($T-\tau$) combinations, respectively. The horizons are indicated in the rows, the maturities in the columns. For the in-sample analysis, we estimate the models using monthly US data from January 1952 to December 2012. In the out-of-sample analysis, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t.

Table 8: Macroeconomic Uncertainty and Economic Value of Bond Excess Return Forecasts

Panel A: Very High Uncertainty

		Extended vs. Standard Estimation										Extended Estimation vs. EH Forecast									
		Uncertainty 1m					Uncertainty 12m					Uncertainty 1m					Uncertainty 12m				
	$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ	1m	-7	42	130	293	259	-17	37	186	341	322	-41	-15	295	436	557	-58	-32	283	384	691
	3m	-16	42	399	468	999	-16	35	485	612	1177	-13	9	172	75	538	-15	8	271	235	546
	12m	11	56	222	424	743	10	55	228	429	773	1	16	152	252	252	2	23	165	243	244
	24m	-1	5	83	158	294	-4	0	68	152	342	-9	-20	51	93	103	-9	-21	52	75	94
	60m	3	11	64	117	187	3	13	67	111	149	4	13	41	56	79	4	19	63	86	74

Panel B: High Uncertainty

		Extended vs. Standard Estimation										Extended Estimation vs. EH Forecast									
		Uncertainty 1m					Uncertainty 12m					Uncertainty 1m					Uncertainty 12m				
	$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ	1m	-5	33	256	380	449	-15	14	85	154	194	-30	-27	171	240	133	-39	-40	67	112	-12
	3m	-2	50	378	494	908	-9	20	241	323	674	-11	-0	180	194	365	-14	-8	90	80	213
	12m	14	57	230	395	677	5	31	135	268	485	1	17	89	126	204	-3	4	39	48	61
	24m	6	25	117	222	408	3	13	68	149	307	-2	-4	21	75	99	-3	-7	4	21	33
	60m	2	5	43	82	176	1	4	36	73	132	1	5	14	6	-15	1	3	18	16	-10

Panel C: Low Uncertainty

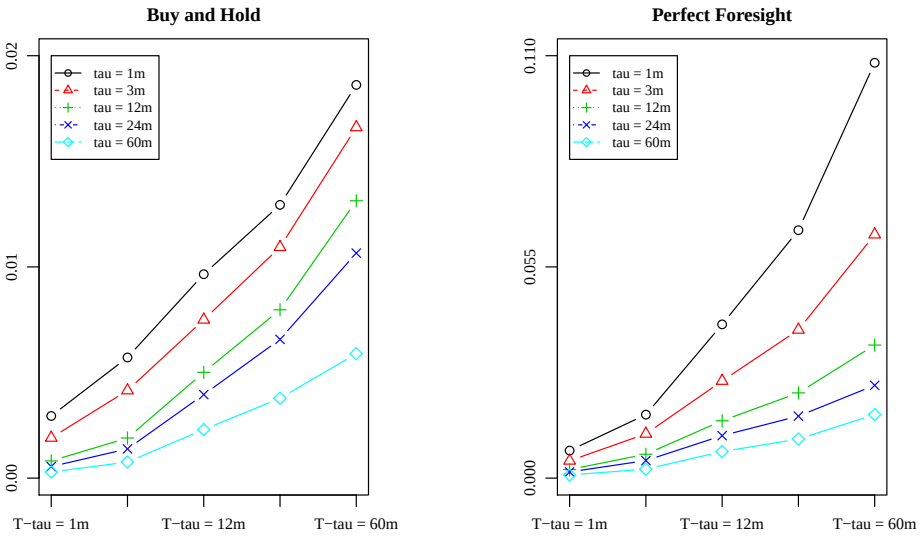
		Extended vs. Standard Estimation										Extended Estimation vs. EH Forecast									
		Uncertainty 1m					Uncertainty 12m					Uncertainty 1m					Uncertainty 12m				
	$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ	1m	-8	-13	3	-45	40	-4	-12	17	-38	199	-10	-21	-18	-30	-115	-7	-18	-21	-42	-70
	3m	-3	-1	42	22	98	-4	-9	1	31	195	-6	-10	4	-16	-130	-5	-13	4	32	37
	12m	4	14	63	95	141	3	10	62	103	212	1	5	28	19	-103	2	11	69	71	-105
	24m	1	4	23	63	201	2	6	43	133	362	1	5	15	1	-12	2	7	7	-12	-40
	60m	1	5	22	52	164	2	6	38	87	263	1	-0	-12	-24	-18	1	1	-6	-21	-39

Panel D: Very Low Uncertainty

		Extended vs. Standard Estimation										Extended Estimation vs. EH Forecast									
		Uncertainty 1m					Uncertainty 12m					Uncertainty 1m					Uncertainty 12m				
	$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ	1m	-5	-16	0	15	104	-6	-2	29	-38	127	-7	-21	-37	-22	-6	-6	-2	15	53	-162
	3m	-6	-14	-48	-89	-87	-2	-3	19	41	178	-6	-14	-27	-15	68	-2	-3	35	70	-35
	12m	-2	-11	-42	-71	-50	3	14	66	103	185	-2	-3	-2	5	-112	3	14	52	-8	-187
	24m	0	-0	2	31	174	-1	-0	9	73	373	0	-0	5	17	-14	-1	-0	5	27	0
	60m	2	6	25	63	216	3	12	78	162	430	1	0	-16	-43	-90	3	0	-1	0	0

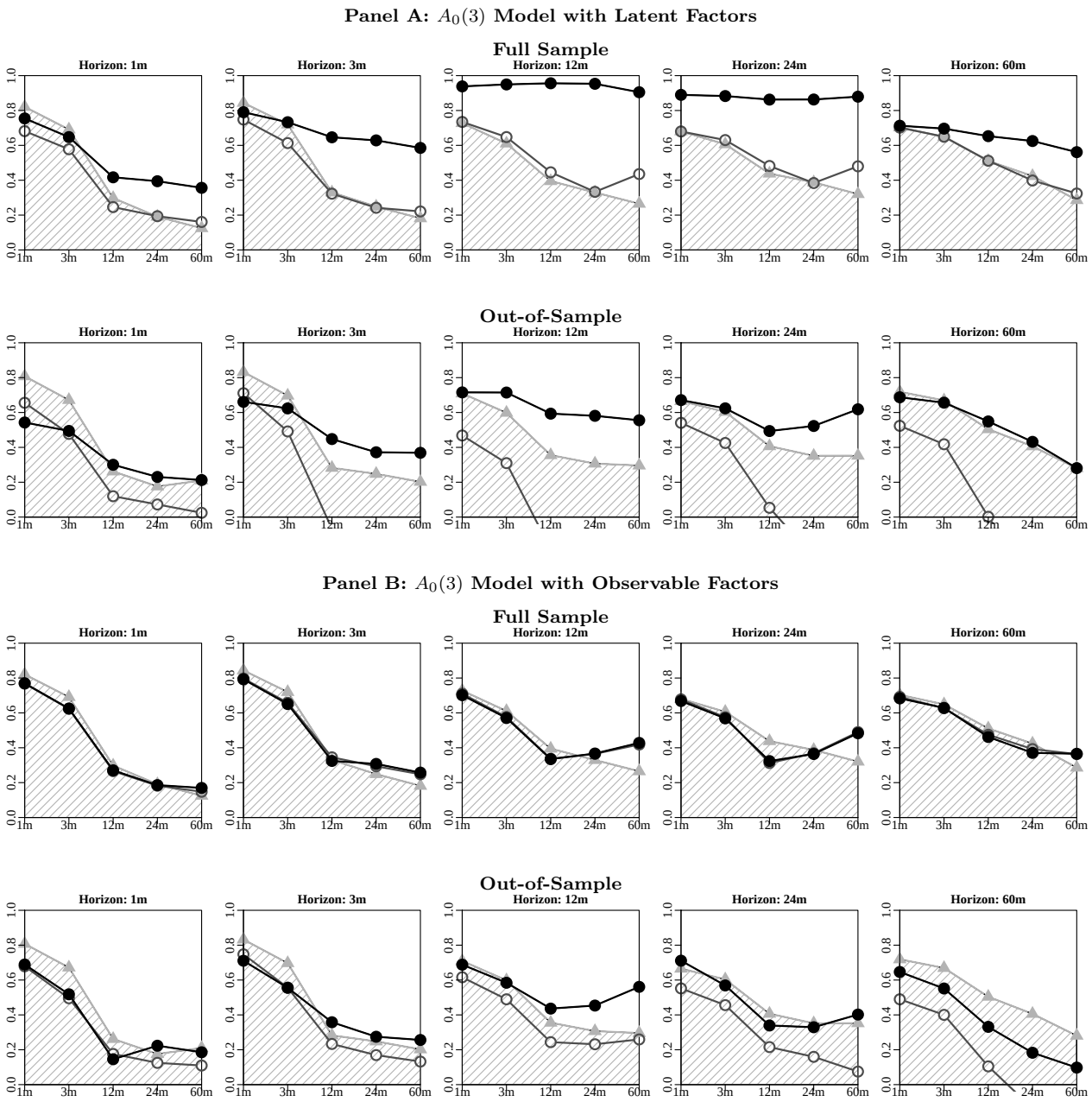
Notes: The Table presents results on how the economic value generated by ATSM forecasts using the extended estimation procedure relates to macroeconomic uncertainty. We measure uncertainty using the data of Jurado et al. (2015) and classify our sample into periods of very high, high, low, and very low uncertainty using the 90%, 75%, 25%, and 10% quantiles of their one-month and twelve-month uncertainty measures. In Panels A to D, we report out-of-sample results for these different levels of uncertainty for 25 horizon (τ) and maturity ($T-\tau$) combinations, with horizons indicated in the rows and maturities in the columns. For an investment horizon τ, an investor with target volatility of $\sigma^* = 2\%$ and relative risk aversion $\rho = 3$ optimally allocates his wealth between the risk-free bond with maturity τ and the risky bond with maturity $T > \tau$. In each Panel, we report the economic value added by the extended compared to the standard estimation and compared to the EH-consistent constant risk premium forecasts using the performance measure of Goetzmann et al. (2007). All performance measures are annualized and reported in basis points. Using monthly US data from January 1952 to December 2012, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t.

Figure 1: Average Excess Returns



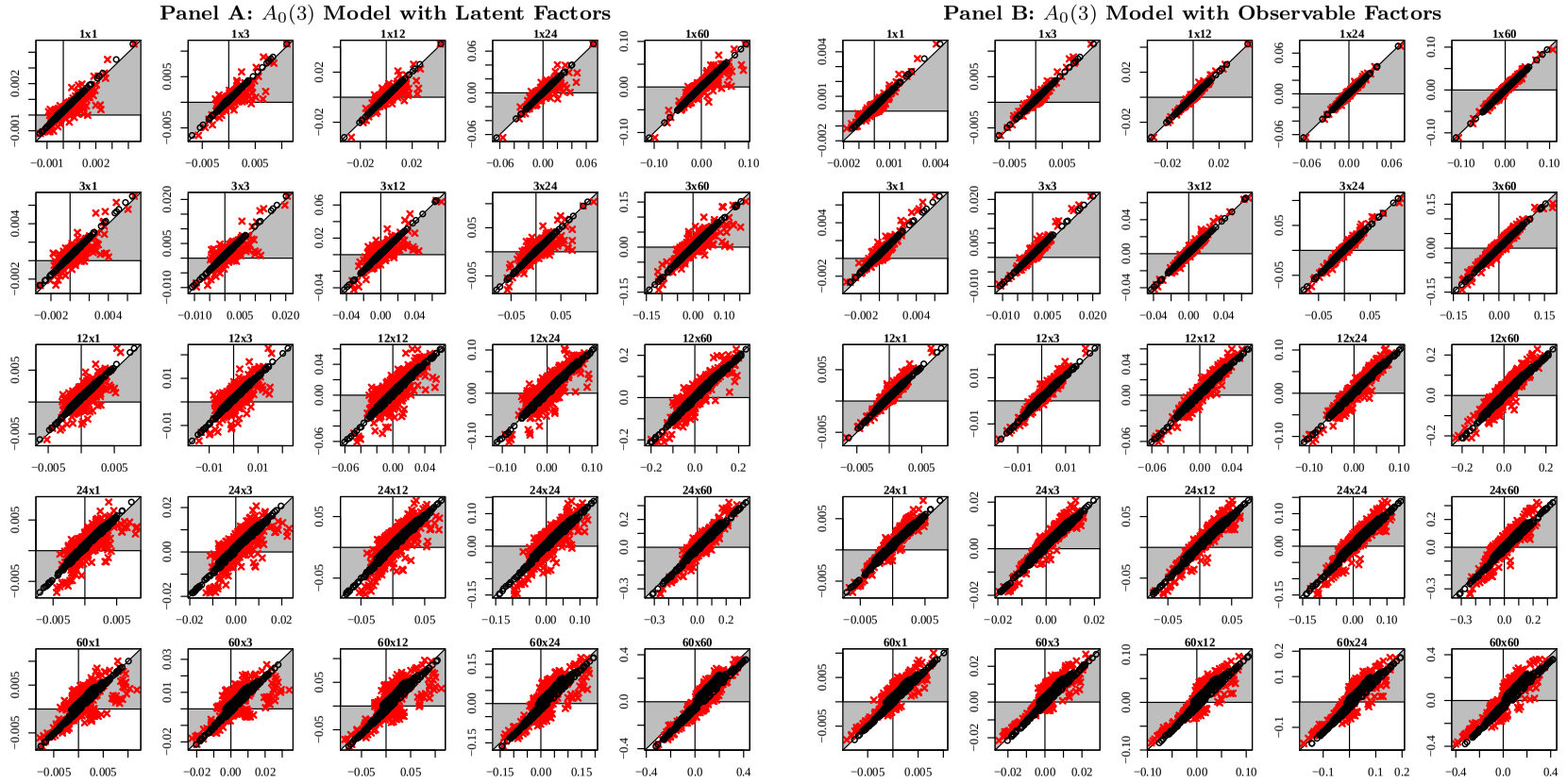
Notes: The Figure plots averages of annualized excess returns of trading the longer-term bond with maturity T over a horizon τ. Each line represents the term structure of excess returns for a given horizon τ for maturities indicated on the x-axis. The left column represents the excess returns of a buy and hold strategy. The right columns plots excess returns of optimal bond portfolios of investors that have perfect foresight; for an investment horizon τ, an investor with target volatility of $\sigma^* = 2\%$ and relative risk aversion $\rho = 3$ optimally allocates his wealth between the risk-free bond with maturity τ and the risky bond with maturity $T > \tau$. The graphs are based on monthly US data from January 1952 to December 2012.

Figure 2: Excess Returns relative to Perfect Foresight



Notes: The Figure plots average excess returns of bond portfolios constructed based on EH and model forecasts relative to average excess returns of portfolios using perfect foresight forecasts. Panel A contains results for the latent factor model, Panel B for the observable factor model. Portfolios using EH-consistent constant risk premium forecasts are represented by the light gray lines (with triangles and shaded area below). Portfolios based on model forecasts using the standard estimation procedure are plotted in dark gray (with circles) and portfolios using extended estimation forecasts are displayed in black (with bullets). The underlying portfolios are generated by an investor with target volatility of $\sigma^* = 2\%$ and relative risk aversion $\rho = 3$ who optimally allocates his wealth between the risk-free bond with maturity τ and the risky bond with maturity $T > \tau$. For each country, we present separate graphs for five horizons τ, where the corresponding maturities are given on the x- axis that displays $T - \tau$. For the in-sample analysis, we estimate the models using monthly US data January 1952 to December 2012. In the out-of-sample analysis, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t.

Figure 3: Forecast Errors and Excess Returns



Notes: This Figure illustrates the relation between out-of-sample forecast errors (x-axis) and realized bond excess returns (y-axis) for the latent and observable factor models in Panels A and B, respectively. Forecast errors from using EH-consistent constant risk premium forecast errors are represented by black circles and forecast errors from using extended estimation ATSM forecasts are represented by red crosses. Shaded areas represent areas where forecasts have the wrong sign and hence forecast errors that lead to (bond portfolio) losses. Each plot represents one of the 25 horizon (τ) and maturity ($T-\tau$) combinations as indicated by the plot labels "$\tau \times (T-\tau)$". We use the long data set for the United States covering monthly data from January 1952 through December 2012. We estimate the first model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t.

Internet Appendix for

The Economic Value of Predicting Bond Risk Premia

(not for publication)

In this separate Internet Appendix, we describe details of the modeling approaches that we use in the paper and report additional empirical results that corroborate our findings.

In Section A, we present details of the specification and estimation of the latent and the observable factor model and describe the factor rotation procedure that we apply to make the estimates of both model parameterizations comparable. We then present the model estimates and provide an interpretation of the (rotated) state variables as well as the drivers of bond risk premia in Section B. In Section C, we derive the optimal portfolio weights for power utility investors. Section D provides evidence that our conclusions are unaffected by the recent financial crisis and Section Section E shows that findings are very similar or even more pronounced for Japan, Switzerland, Germany, and the UK. Finally, we summarize in Section F some additional robustness checks related to the ATSM specification and related to using (combinations of) forward rates as predictors for bond excess returns.

A Details on Model Specification and Estimation

In this Appendix we introduce the specific parameterization of the generic affine model described in Section 2.1.1. The latent-state model is described in Section A.1 below. We also estimate an observable-state Gaussian model according to the Joslin et al. (2011) parameterization, described in A.2.

A.1 Latent-State Model and Estimation

Our base model is a member of the canonical affine $A_0(3)$ class with latent factors. We parameterize for $\mathbb{M} \in \{\mathbb{P}, \mathbb{Q}\}$

$$b^{\mathbb{P}} = \begin{pmatrix} b_1^{\mathbb{P}} \\ b_2^{\mathbb{P}} \\ b_3^{\mathbb{P}} \end{pmatrix}, \beta^{\mathbb{P}} = \begin{pmatrix} \beta_{11}^{\mathbb{P}} & 0 & 0 \\ \beta_{21}^{\mathbb{P}} & \beta_{22}^{\mathbb{P}} & 0 \\ \beta_{31}^{\mathbb{P}} & \beta_{32}^{\mathbb{P}} & \beta_{33}^{\mathbb{P}} \end{pmatrix}, \sigma(X_t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \delta_X = \begin{pmatrix} \delta_{X1} \\ \delta_{X2} \\ \delta_{X3} \end{pmatrix} \tag{A.1}$$

with $r_t = \delta_0 + \delta_X^\top X_t$ and $b_i^{\mathbb{Q}} = 0$, $i = 1, 2, 3$ (the unconditional mean of the short rate is determined by δ_0 under $\mathbb{Q}$).

A.1.1 Standard Estimation Procedure

The transition density $p(X \mid \theta^{\mathbb{P}}, X_{t_0})$ for the entire sample is Gaussian which is known explicitly and in closed form for the $A_0(3)$ model.[15] For a window $[t_m, t_n], t_1 \leq t_m < t_n \leq t_N$ it can be written as

$$p(X_{t_m}, X_{t_{m+1}}, \ldots, X_{t_n} \mid \theta, X_{t_0}) = \prod_{k=m}^{n} p(X_{t_k} \mid X_{t_{k-1}}, \theta^{\mathbb{P}}) \tag{A.2}$$

owing to the Markov property of the process. This is the first expression appearing in the likelihood Eq. (18). The second expression is from Eq. (17)

$$p(y_{t_m}^{T_1}, y_{t_m}^{T_2}, \ldots, y_{t_m}^{T_{24}}, \ldots, y_{t_n}^{T_1}, y_{t_n}^{T_2}, \ldots, y_{t_n}^{T_{24}} \mid X, \theta) \tag{A.3}$$

$$= p(y_{t_m}^{T_1}, y_{t_m}^{T_2}, \ldots, y_{t_m}^{T_24}, \ldots, y_{t_n}^{T_1}, y_{t_n}^{T_2}, \ldots, y_{t_n}^{T_{24}} \mid X_{t_m}, X_{t_{m+1}}, \ldots, X_{t_n}, \theta_{\mathbb{Q}}, \theta_\varepsilon) \tag{A.4}$$

$$= \prod_{k=n}^{m} \prod_{i=1}^{24} \varphi(\varepsilon_{t_k}^{T_i}; 0, \exp(-(c_0 + c_1 T_i + c_2 T_i^2))), \tag{A.5}$$

[15] In the robustness analysis where we estimate $A_1(3)$ models, we use second-order likelihood expansions from Filipović et al. (2013).

where $\varphi(\cdot;\mu,\sigma)$ denotes the normal density for a normally distributed random variable with mean μ and standard deviation σ. In what follows we introduce the short-hand notation $X_{[m,n]} := X_{t_m}, X_{t_{m+1}}, \ldots, X_{t_n}$ and $y_{[m,n]} := y_{t_m}^{T_1}, y_{t_m}^{T_2}, \ldots, y_{t_m}^{T_24}, \ldots, y_{t_n}^{T_1}, y_{t_n}^{T_2}, \ldots, y_{t_n}^{T_24}$.

According to Bayes theorem we have

$$\begin{aligned} p(\theta^{\mathbb{P}} \mid \theta_{\setminus\theta^{\mathbb{P}}}, X_{[m,n]}, y_{[m,n]}) &\propto p(y_{[m,n]} \mid \theta, X_{[m,n]}) p(\theta^{\mathbb{P}} \mid \theta, X_{[m,n]}) \\ &\propto p(X_{[m,n]} \mid \theta) \pi(\theta^{\mathbb{P}} \mid \theta_{\setminus\theta^{\mathbb{P}}}), \end{aligned} \tag{A.6}$$

where $\theta_{\setminus\theta^{\mathbb{P}}}$ denotes the parameter vector θ without $\theta^{\mathbb{P}}$. We use the uninformative prior Eq. (19) for π above. We sample from Eq. (A.6) using Random Walk Metropolis-Hastings (RWMH) steps. The other parameters we sample from

$$\begin{aligned} p(\theta_j \mid \theta_{\setminus\theta^{\mathbb{P}}\cup\theta_j}, X_{[m,n]}, y_{[m,n]}) &\propto p(y_{[m,n]} \mid \theta, X_{[m,n]}) p(\theta_j \mid \theta_{\setminus\theta^{\mathbb{P}}\cup\theta_j}, X_{[m,n]}) \\ &\propto p(y_{[m,n]} \mid \theta, X_{[m,n]}) p(X_{[m,n]} \mid \theta), \end{aligned} \tag{A.7}$$

which we also sample from using RWMH steps. The latent processes are sampled one at a time for $m \leq i \leq n-1$

$$p(X_{t_i} \mid X, \theta, y) \propto p(y_{t_i} \mid X_{t_i}, \theta) p(X_{t_i} \mid X_{t_{i-1}}, \theta) p(X_{t_{i+1}} \mid X_t, \theta). \tag{A.8}$$

For $i = n$, we sample from

$$p(X_{t_n} \mid X, \theta, y) \propto p(y_{t_n} \mid X_{t_n}, \theta) p(X_{t_n} \mid X_{t_{n-1}}, \theta). \tag{A.9}$$

A.1.2 Extended Estimation Procedure

The extended estimation changes from the procedure introduced in Appendix A.1.1 accounting for the additional forecasting equations in Eq. (21). We have the density for the forecast errors again for a sub sample for the time period $[t_m, t_n]$

$$\begin{aligned} h(m,n) &:= \prod_{m\leq k\leq n, 1\leq i\leq 5, 1\leq j\leq J_i} p(\epsilon_{t_k}^{\tau_i,T_{i,j}} \mid \theta_{\epsilon\varepsilon}) \mathbb{1}_{\{t_k+\tau_i\leq t_n\}} \\ &= \prod_{m\leq k\leq n, 1\leq i\leq 5, 1\leq j\leq J_i} \phi(\epsilon_{t_k}^{\tau_i,T_{i,j}}; 0, e^{-(D(c_0+c_1T_{i,j}+c_2T_{i,j}^2))+(d_0+d_1\tau_i+d_2\tau_i^2)}) \mathbb{1}_{\{t_k+\tau_i\leq t_n\}} \end{aligned} \tag{A.10}$$

The density Eq. (A.6) above becomes

$$p(X_{[m,n]} \mid \theta)\pi(\theta^{\mathbb{P}} \mid \theta_{\setminus\theta^{\mathbb{P}}})h(m,n), \tag{A.11}$$

while Eq. (A.7) becomes

$$p(y_{[m,n]} \mid \theta, X_{[m,n]})p(X_{[m,n]} \mid \theta)h(m,n), \tag{A.12}$$

and the density for the latent processes (corresponding to Eq. (A.8)) becomes

$$p(y_{t_i} \mid X_{t_i}, \theta)p(X_{t_i} \mid X_{t_{i-1}}, \theta)p(X_{t_{i+1}} \mid X_t, \theta)h(m,n),$$

respectively

$$p(y_{t_n} \mid X_{t_n}, \theta)p(X_{t_n} \mid X_{t_{n-1}}, \theta)h(m,n),$$

for Eq. (A.9). This modified density ensures that past forecast errors influence the density of the latent states for $i < n$. For the state variables at time t_n nothing can be learned from $h(m,n)$,

since $\epsilon_{t_n}^{\tau_i,T_{i,j}}$ originate from $X_{t_i}, i < n$, so that the estimation remains truly out-of-sample.

A.2 Observable-State Model and Estimation

In this Section we adapt the discrete-time model from Joslin et al. (2011, JSZ) to a continuous-time setting and describe the specific parameterization. Starting from an annualized yield panel $y_1, \ldots, y_n$ we compute the principal components $\mathcal{P}_t := \underbrace{W_{1,n}}_{3\times 24} y_t$ and specify

$$\begin{aligned} d\mathcal{P}_t &= (b^{\mathbb{P}} - \beta^{\mathbb{P}}\mathcal{P}_t)dt + \sigma dW_t^{\mathbb{P}}, \\ d\mathcal{P}_t &= (b^{\mathbb{Q}} - \beta^{\mathbb{Q}}\mathcal{P}_t)dt + \sigma dW_t^{\mathbb{Q}}, \\ r_t &= \rho_{0\mathcal{P}} + \rho_{1\mathcal{P}}\mathcal{P}_t \end{aligned} \tag{A.13}$$

With a completely Gaussian system we can solve this process and write for a discrete time step $\Delta := t_i - t_{i-1}$ for $\mathbb{M} \in \{\mathbb{P}, \mathbb{Q}\}$

$$\Delta\mathcal{P}_t = K_{0\mathcal{P}}^{\mathbb{M}} + K_{1\mathcal{P}}^{\mathbb{M}}\mathcal{P}_{t_{i-1}} + \Sigma_{\mathcal{P}}\epsilon_{t_i}^{\mathbb{M}}, \tag{A.14}$$

where $\epsilon^{\mathbb{M}} \sim N(0, I_3)$, and $K_{0\mathcal{P}}^{\mathbb{M}}, K_{1\mathcal{P}}^{\mathbb{M}}$, and $\Sigma_{\mathcal{P}}$ can be solved explicitly through computing a matrix exponential (Filipović et al., 2013) as a function of $b^{\mathbb{M}}, \beta^{\mathbb{M}}$, and σ. The two representations are related one-to-one, and Eq. (A.14) corresponds exactly to Eqs. (6) and (7) in JSZ. Specifically we set

$$b^{\mathbb{P}} = \begin{pmatrix} b_1^{\mathbb{P}} \\ b_2^{\mathbb{P}} \\ b_3^{\mathbb{P}} \end{pmatrix}, \beta^{\mathbb{P}} = \begin{pmatrix} \beta_{11}^{\mathbb{P}} & 0 & 0 \\ \beta_{21}^{\mathbb{P}} & \beta_{22}^{\mathbb{P}} & 0 \\ \beta_{31}^{\mathbb{P}} & \beta_{32}^{\mathbb{P}} & \beta_{33}^{\mathbb{P}} \end{pmatrix}, \sigma = \begin{pmatrix} s_{11} & 0 & 0 \\ s_{21} & s_{22} & 0 \\ s_{31} & s_{32} & s_{33} \end{pmatrix}. \tag{A.15}$$

The main workhorse of JSZ is then the parameterization of the yield curve indirectly through the parameters $(\underbrace{\lambda^{\mathbb{Q}}}_{3\times1}, \underbrace{k_{\infty}^{\mathbb{Q}}}_{\text{scalar}}, \underbrace{\sigma}_{3\times3})$, which are defined below. To engineer this we follow their algorithm to the letter. The latent-factor $A_0(3)$ model and the observable-factor model are related through (from Eq. (8)) $\mathcal{P}_{t_n} = A_{W_{1,n}} + B_{W_{1,n}} X_t$, where $A_{W_{1,n}} := A_{W_{1,n}} \phi^{\mathbb{Q}}$, and $B_{W_{1,n}} := A_{W_{1,n}} \psi^{\mathbb{Q}}$ with $\phi^{\mathbb{Q}} = (-\phi(T_1)/T_1, \ldots, -\phi(T_{24})/T_{24})^{\top}$ and likewise the matrix $\psi^{\mathbb{Q}} = (-\psi^{\top}(T_1)/T_1, \ldots, -\psi^{\top}(T_{24})/T_{24})^{\top}$. Defining the diagonal matrix $J(\lambda^{\mathbb{Q}}) := \text{diag}(\lambda_1^{\mathbb{Q}}, \lambda_2^{\mathbb{Q}}, \lambda_3^{\mathbb{Q}})$ we first solve the differential equations (6) by setting $\beta^{\mathbb{Q}} = J(\lambda^{\mathbb{Q}})$, and $\rho_{1\mathcal{P}} = 1_3$ (where 1_3 is a column vector with three elements, each of which is equal to one), multiply them with $W_{1,n}$ and denote the resulting matrix by B. Thereafter, we solve the differential equations (7) and (6) by setting $b^{\mathbb{Q}} = k_{\infty}^{\mathbb{Q}} e_1$, where e_1 is the unit vector, $\beta^{\mathbb{Q}} = J(\lambda^{\mathbb{Q}})$, and $\sigma^{\star}(\mathcal{P}_t) = B^{-1}\sigma$, multiply the vector with $W_{1,n}$ and denote it by A. Thereafter we set $\beta^{\mathbb{Q}} = BJ(\lambda^{\mathbb{Q}} B^{-1}$, $b^{\mathbb{Q}} = k_{\infty}^{\mathbb{Q}} B e_{e_1} - \beta^{\mathbb{Q}} A$, $\rho_{1\mathcal{P}} = (B^{-1})^{\top} 1_3$, and $\rho_{0\mathcal{P}} = -A\rho_{1\mathcal{P}}$. This parameterization ensures that the system of principal components as driving factors is consistent with no-arbitrage and the yield panel. Note that we have the same numbers of parameters (10) under $\mathbb{Q}$ in the JSZ parameterization compared to the latent factor model introduced in Section A.1.

The assumption that the principal components of the yield panel are observed without error simplifies the estimation process considerably. The likelihood from Eq. (18) stays the same with X_t being replaced by $\mathcal{P}_t$ and without the prior π. This means that the latent states themselves are not free parameters anymore, and the parameters can be estimated by a simple likelihood search. As noted in (JSZ, Eq. (20)) in this case the parameters $K_{0\mathcal{P}}^{\mathbb{P}}$ and $K_{1\mathcal{P}}^{\mathbb{P}}$ can be estimated independently of $\Sigma_{\mathcal{P}}$ by OLS, equivalently to maximum likelihood. In our case, since $b^{\mathbb{P}}$ and $\beta^{\mathbb{P}}$ are related one-to-one through a nonlinear mapping, this means that we can estimate $b^{\mathbb{P}}$ and $\beta^{\mathbb{P}}$ by nonlinear least squares (NLLS). Subsequently, the parameters $(\lambda^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \sigma)$ are found through maximizing the likelihood function. For the extended estimation we consider

the likelihood function in Eq. (20), again without the prior π. Here we can not determine $b^{\mathbb{P}}$ and $\beta^{\mathbb{P}}$ by NLLS, since the forecasting equation depends on all the parameters simultaneously. For the out-of-sample procedure we pursue the following algorithm

1. At time t_n compute $W_{1,n}$ from the principal components
2. Estimate the parameters either by maximizing Eq. (18) (standard estimation), or Eq. (20) (extended estimation), both without the prior π. Call the parameters $\theta_{1,n}$
3. Make the forecast using $\mathcal{P}_{t_n}$, and $\theta_{1,n}$
4. Increment n by one and go to 1)

A.3 Factor Rotation and Parameter Estimates

The specification of the main model in Section A.1 is in terms of latent factors, while the specification in Section A.2 is in terms of the principal components of the yield curve. To make the factors comparable and interpretable, we rotate both parameterizations into a system where the first factor is the short rate, the second factor is the instantaneous slope (the drift of the short rate), and the third factor is the curvature of the yield curve (the drift of the slope). This rotation is proposed in Collin-Dufresne et al. (2008). The rotation is the same for both parameterizations, we therefore start from the generic Gaussian three-factor model Eq. (4), where the factors can either be latent or observable. For $\mathbb{M} \in \{\mathbb{P}, \mathbb{Q}\}$ we have

$$dX_t = (b^{\mathbb{M}} - \beta^{\mathbb{M}} X_t)dt + \sigma dW_t^{\mathbb{M}},$$

where the short rate is given by $r_t = \delta_0 + \delta_X^\top X_t$. The first factor is the short rate and its dynamics are given by

$$dr_t = \delta_X^\top dX_t = (\delta_X^\top b^{\mathbb{M}} - \delta_X^\top \beta^{\mathbb{M}} X_t)dt + \delta_X^\top \sigma dW_t^{\mathbb{M}}.$$

From this is follows that the slope is $\mu_{1t}^{\mathbb{Q}} := \delta_X^\top b^{\mathbb{Q}} - \delta_X^\top \beta^{\mathbb{Q}} X_t$ with dynamics

$$d\mu_{1t}^{\mathbb{Q}} = (-\delta_X^\top \beta^{\mathbb{Q}} b^{\mathbb{M}} + \delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{M}} X_t)dt - \delta_X^\top \beta^{\mathbb{Q}} \sigma dW_t^{\mathbb{M}}.$$

Finally, the curvature is $\mu_{2t}^{\mathbb{Q}} := -\delta_X^\top \beta^{\mathbb{Q}} b^{\mathbb{Q}} + \delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{Q}} X_t$ with dynamics

$$d\mu_{2t}^{\mathbb{Q}} = (\delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{Q}} b^{\mathbb{M}} - \delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{Q}} \beta^{\mathbb{M}} X_t)dt + \delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{Q}} \sigma dW_t^{\mathbb{M}}.$$

The observables are therefore related through an affine transformation $Y_t := (r_t, \mu_{1t}, \mu_{2t})^\top = \eta + \iota X_t$ to the original states X_t. Inserting the observables into the drift of the SDE above we get the system

$$d\begin{pmatrix} r_t \\ \mu_{1t}^{\mathbb{Q}} \\ \mu_{2t}^{\mathbb{Q}} \end{pmatrix} = \begin{pmatrix} \mu_{1t}^{\mathbb{Q}} \\ \mu_{2t}^{\mathbb{Q}} \\ \gamma + \kappa_0 r_t + \kappa_1 \mu_{1t} + \kappa_2 \mu_{2t} \end{pmatrix} dt + \delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{Q}} \sigma dW_t^{\mathbb{Q}},$$

where we use the notation from Collin-Dufresne et al. (2008), and $\gamma, \kappa_0, \kappa_1, \kappa_2$ are functions of $\delta_0, \delta_X, \beta^{\mathbb{Q}}$, and $b^{\mathbb{Q}}$. Under $\mathbb{P}$

$$dY_t = (\kappa\theta^{\mathbb{P}} + \kappa^{\mathbb{P}} Y_t)dt + \delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{Q}} \sigma dW_t^{\mathbb{P}},$$

where

$$\kappa\theta^{\mathbb{P}} = \begin{pmatrix} \delta_X^\top b^{\mathbb{P}} \\ -\delta_X^\top \beta^{\mathbb{Q}} b^{\mathbb{P}} \\ \delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{Q}} b^{\mathbb{P}} \end{pmatrix} - \begin{pmatrix} -\delta_X^\top \beta^{\mathbb{P}} \\ \delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{P}} \\ -\delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{Q}} \beta^{\mathbb{P}} \end{pmatrix} \iota^{-1}\eta,\ \kappa^{\mathbb{P}} = \begin{pmatrix} -\delta_X^\top \beta^{\mathbb{P}} \\ \delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{P}} \\ -\delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{Q}} \beta^{\mathbb{P}} \end{pmatrix} \iota^{-1}.$$

In Table IA.1 we report the parameters $\gamma, \kappa_0, \kappa_1, \kappa_2$ and the lower triangle of the instantaneous covariance matrix $V = \delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{Q}} \sigma (\delta_X^\top \beta^{\mathbb{Q}} \beta^{\mathbb{Q}} \sigma)^\top$, as well as the matrices $\kappa\theta^{\mathbb{P}}$ and $\kappa^{\mathbb{P}}$.

B Additional Estimation Results

B.1 Parameter Estimates

We report parameter estimates of the rotated models of the latent-factor specification (Section A.1) and the Joslin et al. (2011, JSZ) specification (Section A.2) in Table IA.1. Differences between the parameter estimates stem from the estimation technology (Bayesian vs. Maximum Likelihood vs. standard vs. extended estimation), and from the assumption behind the JSZ formulation in terms of observable factors that a weighted sum of yields, the principal components, are observed without error.

It can be seen that the parameter estimates from the JSZ specification reported in Panel A look similar between standard and extended estimation. Here, the states are fixed and observable, and thus remain unchanged even when taking into consideration past forecast errors. Comparing this to the estimates of the latent factor specification, where the estimates of the factors do change between standard and extended estimation procedure, we interpret this as an indication that the information of past forecast errors is largely reflected in the estimates of the latent state variables. Finally, the assumption of a yield portfolio observed without error behind the observable factors of JSZ explains the differences between JSZ $\mathbb{Q}$ parameter

estimates and those of the latent-factor model standard estimation. The yield pricing errors of the latent-factor model standard estimation are greater on average at the short end than the JSZ pricing errors (the first observable factor resembles the short rate), but smaller on average at the long end.

B.2 Interpretation of Rotated State Variables and Drivers of Bond Risk Premia

Given the estimates from the system of rotated factors, we now attach an economic interpretation to the state variables and to the drivers of bond risk premia. Conceptually, the rotated state variables Y obtained through the procedure just described in Section A.3 represent the model-implied (instantaneous) short rate level, slope, and curvature. In what follows, we discuss the relation between these model-implied quantities to model-free estimates of level, slope, and curvature that we compute from the principal components of yields following Collin-Dufresne et al. (2008). Our results suggest that the model states are strongly related to level, slope, and curvature and/or combinations thereof and that the cross-section of bond risk premia can be characterized in terms of these economically interpretable characteristics.

Table IA.2 presents correlations of the model-implied state variables Y with the corresponding model-free estimates. For the observable factor model (Panel B), we find that the model states Y_1, Y_2, and Y_3 exhibit high correlations with model-free level, slope, and curvature of 99.9%, 97.9%, and 75.3%, respectively. Furthermore, regressing each Y on the three model-free estimates results in R^2s of 99.94% or higher. For the latent factor model in Panel A, we find that the state variables of the standard estimation procedure are strongly related to level (correlation of 99.2%), slope (62.2%), and curvature (37.9%) as well. For the extended estimation,

we find that Y_1 is highly correlated to the short rate level (96.5%) but that it is also exhibits a relatively strong correlation with slope (28.9%) and curvature (-15.6%). Taken together, level, slope, and curvature explain almost 98% of the variation in Y_1. The other two state variables also exhibit strong correlations with level and slope but little with curvature. A large share of their variation can be attributed to level, slope and curvature as well, with regression R^2s of 66% and 68% for Y_2 and Y_3, respectively. Overall, our results suggest that level, slope, and curvature and/or linear combinations thereof are important economic determinants of the model state variables.

To identify the drivers of bond risk premia, we compute their principal components across the 25 horizon/maturity combinations used in the paper. We now focus on the models estimated with the extended procedure because our empirical results reported in the paper show that they match bond excess returns more accurately than the models estimated with the standard procedure. The results reported in Table IA.3 show that bond risk premia from the latent factor model (Panel A) are almost entirely driven by the first principal component (PC_1). PC_1 explains 99% of the variation in bond risk premia and, in turn, the state variables Y explain almost 97% of the variation in PC_1. For the observable factor model (Panel B), we find that PC_1 accounts for 91% of the variation in bond risk premia and that this variation is fully explained by the state variables.[16] Given the correlation structure in Table IA.2, the loadings of the state variables on PC_1 of observable factor model risk premia suggest that risk premia decrease with short rate level and slope whereas they increase with curvature. Given the analogous information for the latent factor model, the loadings on Y_1 suggest the same pattern with respect to level, slope, and curvature but the opposite based on loadings of Y_2 and Y_3. While this finding suggests that there is no one-to-one mapping between the model

[16]The 100% R^2 is not surprising since, both, the observable state variables as well as the model-free estimates of level, slope, and curvature are computed from the principal components of yields.

state variables, risk premia, and empirical measures of level, slope, and curvature, our results nevertheless show that (linear combinations of) level, slope, and curvature are the driving force behind the term structure of bond risk premia.

Overall, these results are consistent with the arguments provided at the outset of the description of the model estimation procedures. The extended estimation of the latent factor model is most flexible to account for predictive information, even if it is not contained in the cross section of yields, thereby improving the model's predictive ability for bond excess returns. While level, slope, curvature and linear combinations thereof explain the largest share of variation in state variables and bond risk premia, the model also picks up information beyond, similar to the linear combination of forward rates by Cochrane and Piazzesi (2005) and in line with the hidden factor of Duffee (2011).

C Power Utility Investors

To ensure that our conclusions about the economic value attainable from (predictable) EH deviations are not specific to assuming mean-variance preferences, we also consider power utility investors in Section 4.3. We now describe the setup for this analysis in detail. For an investment horizon τ, the investor chooses to allocate his wealth between bonds with maturities τ and $T > \tau$. Since the maturity of the shorter-term bond exactly matches the investment horizon, the τ-bond represents the risk-free asset paying a return of y_t^τ. The longer-term T-bond, with remaining maturity $T - \tau$ at the end of the horizon, represents the risky asset. The investor uses model k to generate the conditional expectation of the risky asset return $\mu_{t+\tau}^{T,k}$ which she uses to determine the risky portfolio weight w_t^k.

Power utility investors have constant relative risk aversion (CRRA) ρ and derive utility from

wealth $W_{t+\tau}$ in the form of

$$U(W_{t+\tau}) = \frac{W_{t+\tau}^{1-\rho}}{1-\rho}. \tag{C.16}$$

$W_{t+\tau}$ is determined by initial wealth at time t and the performance of the bond portfolio,

$$W_{t+\tau} = W_t \times R_{t+\tau}^k, \tag{C.17}$$

where $R_{t+\tau}^k = 1 + y_t^\tau + w_t^k \cdot rx_{t+\tau}^T$ is the gross portfolio return. We define the log of wealth $\omega_{t+\tau} \equiv \log(W_{t+\tau})$ and log portfolio return $r_{t+\tau}^k \equiv \log(R_{t+\tau}^k)$ because maximizing the expected log portfolio return allows us to compute the optimal portfolio weights in closed form. The objective function is

$$\max_{w_t^k} \left[y_t^\tau + w_t^k(\mu_{t+\tau}^{T,k} - y_t^\tau) + \frac{1}{2}(1-w_t^k)w_t^k\sigma_t^2 \right] + \frac{1}{2}(1-\rho)w_t^{k^2}\sigma_t^2 \tag{C.18}$$

and the resulting first order condition is given by

$$\mu_{t+\tau}^{T,k} - y_t^\tau + \frac{1}{2}\sigma_t^2 - \rho w_t^k \sigma_t^2, \tag{C.19}$$

which solves for the optimal risky portfolio weight

$$w_t^k = \frac{\mu_{t+\tau}^{T,k} - y_t^\tau + \frac{1}{2}\sigma_t^2}{\rho\sigma_t^2}. \tag{C.20}$$

The weight of the riskless asset is given by $1 - w_t^k$. To measure the economic value generated by the models, we compute the performance measure Θ proposed by Goetzmann et al. (2007), see Eq. (24), and present results in the main text in Section 4.3 and Table 5.

D Recent Financial Crisis

To analyze the benefits of modeling time-varying risk premia during the recent financial crisis, we define two subsamples over the period end of June 2007 to end of December 2012. First, we consider all model forecasts that have corresponding realizations during the crisis (Crisis Subsample 1). Second, we consider only those cases where forecasts and realizations fall into the crisis period (Crisis Subsample 2). We evaluate the in- and-out-of-sample forecasts of the latent and the observable factor models over these two subsamples and report, analogue to the empirical analysis in the paper, measures of forecast ability (Tables IA.7 and IA.4), economic value (Tables IA.8 and IA.5), and directional accuracy (Tables IA.9 and IA.6).

We find, very similar to the full sample period, that the extended estimation procedure dominates the standard estimation procedure statistically (because it has smaller forecast errors as measured by $R2$) and economically (as measured by the Θ estimates following Goetzmann et al. (2007)). The latent factor model beats the EH from a statistical perspective in-sample but not out-of-sample. Moreover, investors cannot benefit from using these model forecasts for bond portfolio allocation decisions instead of presuming that the EH holds. The observable factor model performs worse than EH forecasts in terms of predictive ability and as judged by economic value, both, in- as well as out-of-sample. Furthermore, the directional accuracy of the latent and the observable factor model forecasts appears to be low as compared to using the historical average.

Overall, the evaluation of the models over the crisis subsample corroborates our conclusion that it is difficult for investors to beat the historical average and that the EH can be viewed as a useful benchmark for bond portfolio investments out-of-sample.

E International Evidence

To check that our findings are not specific to the US bond market, we construct data sets of monthly interest rates from money market and swap rates for Switzerland, Germany, the UK, and Japan. Furthermore, we also construct such a data set for the US, which allows for a direct comparison of all five countries using the same kind of rates over the same sample period and additionally allows for a robustness check of our findings based on the long US data over a recent subsample.

Specifically, our data set comprises money market (Libor) rates with maturities of 1 through 11 months, and swap rates with maturities of 1 to 10 years. We bootstrap riskless zero-coupon yields from these money market and swap rates; Feldhütter and Lando (2008) show that swap rates are the best parsimonious proxy for riskless rates.[17] Given the availability of data, our sample period starts in April 1987 for Germany, the UK and the US, January 1988 for Switzerland, and September 1989 for Japan. For each country, we estimate the latent factor model using the standard and the extended estimation procedure.

The conclusions we draw from our international analysis are qualitatively the same as those based on the long US data. Quantitatively the results are even more pronounced, also when comparing the recent subsample of US data to the original sample from 1952 to 2012. Table IA.10 shows that the models fit the term structures of interest rates well but also that there is a tradeoff when jointly fitting yields and past bond excess returns. The standard estimation typically results in smaller yield pricing errors than the extended estimation, in particular at

[17] We have also used bond prices (from Datastream) to directly estimate zero yields and results are qualitatively identical to those reported in this Internet Appendix. Furthermore, we have used term structure data provided by central banks (for countries where data is available) and reach the same conclusions. Thus, our findings do not depend on the mechanism used to estimate the zero curve in general and, more specifically, our conclusions are not affected by credit risk issues that have become relevant in Libor and swap markets during the recent crisis.

the short end of the term structure. Model-implied risk premia from the extended estimation significantly predict bond excess returns with high regression R^2s while this is generally not the case for risk premia estimated with the standard procedure (see Table IA.11). The results in Table IA.12 show that the extended estimation forecasts are more accurate than the standard estimation forecasts and that they beat EH-consistent constant risk premium forecasts in-sample and mostly also out-of-sample. Switching from the standard to the extended estimation provides economic value to mean-variance investors but portfolios based on the model forecasts typically only generate a premium return beyond EH-forecast portfolios in the full sample analysis (see Table IA.13).

Thus, in line with our findings in the paper, the extended dominates the standard estimation procedure in terms of forecast accuracy and economic value attainable from bond investments. The extent to which the validity of the EH is challenged seemingly depends on whether statistical or economic metrics are applied and the apparent conflicts in results can be resolved based on the arguments that we provide in Section 5.2 of the paper. On the one hand, the model forecasts are statistically accurate in terms of squared pricing errors but less so in terms of directional accuracy (Table IA.14). On the other hand, Figures IA.1 and IA.2 show that there is not much potential to beat the EH, in particular for long-term bond investments. Portfolios based on constant risk premium forecasts already capture a large fraction of perfect foresight returns and EH deviations are economically important mainly at short horizons.

F Summary of other Robustness Checks

Alternative ATSM Specifications We verify that our conclusions are robust to changes in the ATSM specification and repeat the empirical analysis using a larger model with four

factors ($A_0(4)$ model) and a stochastic volatility model ($A_1(3)$ model). In general, changing the specification can have an impact on yield pricing errors and/or forecast accuracy. We find that changing the specification may improve or deteriorate particular results but the overall picture does not change and our conclusions remain the same.

Forecasting Bond Excess Returns with Forward Rates Previous research documents (in-sample) predictability of bond excess returns using lagged forward rates; see e.g. Fama and Bliss (1987) and Cochrane and Piazzesi (2005, CP). While we do not impose the CP-factor in the model structure, the extended estimation procedure that matches model risk premia to the data incorporates forward rates that the CP-factor is based upon and it additionally accounts for past forecast errors. In line with previous research, we find that forward rates contain information for in-sample predictions of bond excess returns but do not generate economic value out-of-sample (Thornton and Valente, 2012). Forecasts based on the extended ATSM estimation proposed in this paper have larger predictive ability and add more economic value than the CP-factor forecasts in- and out-of-sample, thus, posing a stronger challenge to the EH and thereby providing more general findings.

Table IA.1: Parameter Estimates

	$A_0(3)$ Model with Latent Factors		$A_0(3)$ Model with ObservableFactors	
	Extended Est.	Standard Est.	Extended Est.	Standard Est.
Panel A: Model Parameters				
γ	0.00241	0.00101	0.00444	0.00437
	(0.00140, 0.00454)	(0.00089, 0.00117)	(0.00416, 0.00483)	(0.00407, 0.00478)
κ_0	−0.05758	−0.01089	−0.03101	−0.03115
	(−0.09608, −0.03488)	(−0.01232, −0.00967)	(−0.03525, −0.02830)	(−0.03463, −0.02784)
κ_1	−2.32112	−0.38277	−1.30344	−1.30828
	(−2.88405, −1.96051)	(−0.40504, −0.35243)	(−1.37531, −1.25959)	(−1.36375, −1.24708)
κ_2	−8.51487	−1.33289	−3.19329	−3.19593
	(−10.41583, −7.13854)	(−1.39376, −1.28505)	(−3.30661, −3.09203)	(−3.29479, −3.07256)
$\kappa\theta_1^{\mathbb{P}}$	0.05993	−0.00600	0.00427	0.00246
	(0.04976, 0.06697)	(−0.01387, 0.00383)	(−0.00055, 0.00715)	(−0.00920, 0.01206)
$\kappa\theta_2^{\mathbb{P}}$	−0.00904	0.00338	0.00282	0.01725
	(−0.01210, −0.00512)	(−0.00424, 0.01422)	(−0.01579, 0.03107)	(−0.02736, 0.04740)
$\kappa\theta_3^{\mathbb{P}}$	−0.01513	0.00236	−0.00416	−0.03452
	(−0.04485, 0.00998)	(−0.01040, 0.01048)	(−0.09203, 0.05986)	(−0.12473, 0.09579)
$\kappa_{11}^{\mathbb{P}}$	−1.76301	−0.00568	−0.31228	−0.14764
	(−1.89751, −1.60112)	(−0.11355, 0.04059)	(−0.38831, −0.25317)	(−0.37978, −0.05065)
$\kappa_{12}^{\mathbb{P}}$	−1.39052	0.70205	0.61176	0.60639
	(−1.75221, −0.83215)	(−0.24718, 1.82774)	(0.34286, 0.95809)	(0.04534, 1.12219)
$\kappa_{13}^{\mathbb{P}}$	−0.25837	0.35922	0.10796	0.10828
	(−0.31572, −0.15812)	(−0.70558, 1.07274)	(0.00634, 0.21368)	(−0.06173, 0.28368)
$\kappa_{21}^{\mathbb{P}}$	0.41839	0.01477	1.00595	0.33529
	(0.37575, 0.47347)	(−0.03189, 0.04808)	(0.67935, 1.49846)	(0.01397, 1.18734)
$\kappa_{22}^{\mathbb{P}}$	0.24030	0.47599	−0.89336	−0.64670
	(0.05570, 0.42758)	(−1.02310, 1.50023)	(−2.89281, 0.46547)	(−3.17144, 1.56808)
$\kappa_{23}^{\mathbb{P}}$	0.06615	1.28280	0.30995	0.35539
	(0.03831, 0.09067)	(0.20429, 2.37349)	(−0.31246, 0.85125)	(−0.51403, 1.03084)
$\kappa_{31}^{\mathbb{P}}$	−0.32749	−0.04836	−2.97870	−1.25927
	(−0.83686, −0.01460)	(−0.08819, 0.00568)	(−4.63063, −1.98418)	(−3.66917, −0.12298)
$\kappa_{32}^{\mathbb{P}}$	−0.99136	−1.59124	0.25473	−0.79998
	(−3.10531, 0.31175)	(−2.82169, 0.14410)	(−4.11336, 6.75652)	(−7.02214, 6.96363)
$\kappa_{33}^{\mathbb{P}}$	−0.24066	−2.39846	−1.80904	−2.01901
	(−0.43551, −0.07278)	(−3.66451, −0.94639)	(−3.50636, 0.19219)	(−3.90864, 0.63201)
V_{00}	0.00031	0.00039	0.00037	0.00037
	(0.00030, 0.00035)	(0.00034, 0.00041)	(0.00034, 0.00042)	(0.00033, 0.00041)
V_{10}	−0.000081	−0.00024	−0.00067	−0.00065
	(−0.000089, −0.000048)	(−0.00029, −0.00019)	(−0.00078, −0.00056)	(−0.00076, −0.00054)
V_{11}	0.00061	0.00053	0.00424	0.00406
	(0.00047, 0.00095)	(0.00044, 0.00063)	(0.00380, 0.00491)	(0.00369, 0.00473)
V_{20}	0.00004	0.00018	0.00170	0.00163
	(−0.00021, 0.000042)	(0.00011, 0.00025)	(0.00136, 0.00207)	(0.00129, 0.00197)
V_{21}	−0.00514	−0.00063	−0.01250	−0.01197
	(−0.00974, −0.00352)	(−0.00077, −0.00050)	(−0.01492, −0.01100)	(−0.01419, −0.01054)
V_{22}	0.04509	0.00080	0.03747	0.03594
	(0.02727, 0.10318)	(0.00062, 0.00101)	(0.03227, 0.04623)	(0.03070, 0.04393)
Panel B: Parameters of the Pricing Error Distribution				
c_0	−4.76626	−5.24472	−6.32887	−6.34020
	(−4.80550, −4.72611)	(−5.28155, −5.17237)	(−6.36477, −6.30660)	(−6.36489, −6.30692)
c_1	−1.18073	−1.29581	−0.43530	−0.43270
	(−1.21364, −1.15620)	(−1.33560, −1.26606)	(−0.45130, −0.41785)	(−0.45141, −0.41822)
c_2	0.11545	0.12412	0.04431	0.04451
	(0.11190, 0.11867)	(0.12041, 0.12713)	(0.04271, 0.04636)	(0.04277, 0.04641)
Panel C: Parameters of the Forecast Error Distribution				
d_0	−4.13905		−2.95676	
	(−4.21057, −4.04543)		(−3.23274, −2.76122)	
d_1	0.44057		1.02304	
	(0.38763, 0.46763)		(0.99223, 1.05790)	
d_2	−0.02519		−0.14332	
	(−0.03035, −0.01555)		(−0.14951, −0.13753)	
D	0.13014		0.28145	
	(0.11703, 0.14379)		(0.23853, 0.30798)	

Notes: This table shows parameter estimates from estimations of the continuous-time Gaussian three factor model under the latent factor parametrization A.1 (left two columns) and the Joslin et al. (2011) parametrization (right two columns) from Section A.2. Panel A shows the estimates of the parameters of the term structure model, Panel B of the observation error parameterization, and Panel C of the forecast error parameterization. The parameters reported in Panel A are in terms of a rotation in short rate, instantaneous slope and instantaneous curvature according to Collin-Dufresne et al. (2008) outlined in Section A.3. We estimate the models using monthly US data from January 1952 to December 2012 using Bayesian MCMC methods (left two columns), and a combination of nonlinear least squares and maximum likelihood (right two columns). The numbers in parentheses are bootstrapped 5%, respectively 95% confidence intervals.

Table IA.2: Interpretation of State Variables

Panel A: $A_0(3)$ Model with Latent Factors

Standard Estimation Procedure

	Y_1	Y_2	Y_3
Level	99.19	−12.56	−3.34
Slope	14.73	62.18	−76.63
Curvature	−8.05	−7.83	37.87
R^2	99.10	73.16	65.71

Extended Estimation Procedure

	Y_1	Y_2	Y_3
Level	96.50	78.20	−80.61
Slope	28.91	19.79	−21.45
Curvature	−15.60	0.91	7.04
R^2	97.76	65.55	67.64

Panel B: $A_0(3)$ Model with Observable Factors

Standard Estimation Procedure

	Y_1	Y_2	Y_3
Level	99.93	24.09	−32.37
Slope	4.61	97.89	−96.28
Curvature	1.15	−62.19	75.31
R^2	100.00	99.95	99.94

Extended Estimation Procedure

	Y_1	Y_2	Y_3
Level	99.94	23.61	−32.03
Slope	4.80	97.96	−96.31
Curvature	1.03	−62.19	75.72
R^2	100.00	99.95	99.94

Notes: The Table presents results for the interpretation of the model state variables. We report pairwise correlations of rotated state variables Y with the CDGJ model-free estimates of instantaneous level, slope, and curvature as well as R^2s from regression Y_i on level, slope, and curvature. Panel A contains results for the latent factor model, Panel B for the observable factor model. We estimate the models using monthly US data from January 1952 to December 2012.

Table IA.3: Drivers of Bond Risk Premia

Panel A: $A_0(3)$ Model with Latent Factors

Standard Estimation Procedure

	PC_1	PC_2	PC_3
Cum. Variation	97.11	99.94	100.00
Y_1	11.09	27.96	−95.23
Y_2	−77.13	−59.12	8.36
Y_3	53.90	79.77	5.72
R^2	99.21	92.83	99.80

Extended Estimation Procedure

	PC_1	PC_2	PC_3
Cum. Variation	99.03	99.91	100.00
Y_1	−37.19	−32.47	−86.83
Y_2	36.29	−67.90	−62.54
Y_3	−37.60	59.49	69.93
R^2	96.60	98.95	99.43

Panel B: $A_0(3)$ Model with Observable Factors

Standard Estimation Procedure

	PC_1	PC_2	PC_3
Cum. Variation	90.53	99.34	100.00
Y_1	−12.02	−5.24	99.14
Y_2	−36.18	−93.16	−3.39
Y_3	14.76	98.63	−7.43
R^2	100.00	100.00	100.00

Extended Estimation Procedure

	PC_1	PC_2	PC_3
Cum. Variation	90.82	98.86	100.00
Y_1	−6.06	−15.74	98.57
Y_2	−35.61	−92.75	−11.36
Y_3	12.86	99.14	2.30
R^2	100.00	100.00	100.00

Notes: The Table presents results for the drivers of bond risk premia. We report pairwise correlations of bond risk premium principal components (PC) and rotated model state variables (Y). The first rows additionally report the cumulative variation explained up to PC_i and, in the last rows, R^2 is from a regression of PC_i on Y. Panel A contains results for the latent factor model, Panel B for the observable factor model. We estimate the models using monthly US data from January 1952 to December 2012.

Table IA.4: Statistical Accuracy of Bond Excess Return Forecasts (Crisis Subsample 1)

Panel A: $A_0(3)$ Model with Latent Factors

	In-Sample										Out-of-Sample									
	Extended vs. Standard Estimation					Model vs. EH Forecast					Extended vs. Standard Estimation					Model vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	−0.08	0.3	0.3	0.31	0.21*	0.28*	0.27*	0.17	0.24	0.19	−0.03	0.09	0.14	0.13*	0.09*	0.31*	0.06	−0.06	−0.06°	−0.02
3m	0.52	0.6	0.41	0.49	0.43*	0.5*	0.45*	0.28	0.42	0.42*	0.15	0.38*	0.26*	0.25*	0.22*	0.1	0.04	−0.16°	−0.12°	−0.06°
12m	0.71	0.63	0.62	0.88	0.79*	0.52	0.43	0.5	0.84	0.8*	0.42*	0.46*	0.55*	0.59*	0.46*	−0.43	−0.31°	−0.14	0.04	−0.1
24m	0.79	0.8	0.84*	0.8*	0.52	0.65	0.7	0.8*	0.77*	0.55*	0.36*	0.4*	0.47*	0.45*	0.3	−0.65°	−0.42	−0.15	−0.12	−0.31
60m	−0.21	−0.22	−0.19	−0.18	−0.31	−0.43	−0.34	−0.26	−0.28	−0.47	0.29	0.27	0.22	0.15	−0.03	−0.58	−0.49	−0.47	−0.54°	−0.89°

Panel B: $A_0(3)$ Model with Observable Factors

	In-Sample										Out-of-Sample									
	Extended vs. Standard Estimation					Model vs. EH Forecast					Extended vs. Standard Estimation					Model vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	−0.08°	−0.06°	−0.02°	0	0.01*	0.15	−0.26	−0.34°	−0.15°	0.01	−0.08°	−0.04°	0.04	0.04*	0.02*	0.1	−0.36	−0.39°	−0.16°	0.01
3m	−0.1°	−0.07°	−0.03°	0	0.02*	−0.22	−0.67	−0.4°	−0.21°	0.02	−0.07°	−0.03°	0.04	0.06*	0.06*	−0.33	−0.8	−0.44°	−0.21°	0.03
12m	−0.05°	−0.05°	−0.05°	−0.03°	0.03*	−0.85	−0.81	−0.75°	−0.55	−0.03	0.04	0.05	0.08	0.1*	0.16*	−0.9	−0.85°	−0.73°	−0.49	0
24m	−0.01	−0.03	−0.04°	−0.03°	0.01	−0.92°	−0.79°	−0.66°	−0.53	−0.16	0.09	0.09	0.09*	0.1*	0.16*	−0.93°	−0.8°	−0.65	−0.51	−0.18
60m	−0.02	−0.02°	−0.03°	−0.03°	−0.01°	−0.64°	−0.56	−0.52	−0.47	−0.42	0.1*	0.09*	0.09*	0.1*	0.13*	−0.87°	−0.78°	−0.74°	−0.7°	−0.71°

Notes: The Table presents $R2$ statistics measuring the predictive ability of the ATSM estimated with the extended procedure relative to the standard estimation and EH-consistent constant risk premium benchmarks. The $R2$ statistic is defined as $R2 \equiv 1 - MSE^m/MSE^b$ where MSE^m and MSE^b denote the mean squared errors of the model and benchmark forecasts, respectively. We report in-sample and out-of-sample results for 25 horizon (τ) and maturity ($T-\tau$) combinations for the latent and observable factor model in Panels A and B, respectively. The horizons are indicated in the rows, the maturities in the columns. * ($^\circ$) identifies $R2$s that are significantly positive (negative) as judged by confidence intervals obtained as the 5%- and 95%-percentiles from a block bootstrap procedure. For the in-sample analysis, we estimate the models using monthly US data from January 1952 to December 2012. In the out-of-sample analysis, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t. We evaluate the statistical accuracy of forecasts that have their corresponding realization occurring during the financial crisis between July 2007 and December 2012.

Table IA.5: Economic Value of Bond Excess Return Forecasts (Crisis Subsample 1)

Panel A: $A_0(3)$ Model with Latent Factors

	In-Sample															Out-of-Sample													
	Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast					Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	9	62	155	251	495	2	40	132	206	335	-6	-3	-0	-0	22	14	62	151	245	376	2	17	112	293	492	-1	-2	-0	-0	-93
3m	21	89	143	237	459	5	43	87	150	346	-1	-2	-0	-0	-6	22	91	139	231	374	9	69	162	314	660	-1	-0	-0	-0	-89
12m	19	57	131	227	440	-0	4	15	86	139	-0	-0	-0	-0	-0	19	57	128	224	372	19	57	183	387	784	-0	-0	-0	-0	-69
24m	18	56	142	233	419	-0	-0	-0	25	74	-0	-0	-0	-0	-0	18	56	140	217	347	19	60	207	382	728	-0	-0	-0	-15	-80
60m	15	46	141	213	358	-0	-0	-0	35	182	-0	-0	-0	-0	-0	15	46	111	11	-359	4	14	164	152	-0	-0	-0	-26	-178	-599

Panel B: $A_0(3)$ Model with Observable Factors

	In-Sample															Out-of-Sample													
	Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast					Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	3	6	-48	37	184	2	-0	-13	-17	-4	-13	-58	-205	-221	-299	-3	-20	-90	44	182	-0	-0	-0	147	67	-18	-85	-244	-207	-295
3m	10	24	-4	40	148	1	-6	-5	9	8	-12	-67	-148	-205	-330	6	12	-15	42	147	-3	10	29	65	56	-17	-79	-156	-197	-327
12m	12	36	70	95	283	-3	-8	-6	-3	9	-7	-21	-62	-135	-165	12	36	48	94	282	1	3	14	25	33	-7	-21	-80	-132	-164
24m	18	56	79	123	279	-0	-0	-11	-0	-0	-0	-0	-62	-112	-146	18	56	69	122	281	3	17	10	8	11	-0	-0	-70	-111	-148
60m	15	46	53	-139	-358	-0	-0	-51	-95	-0	-0	-0	-76	-294	-597	15	46	-92	-186	-359	14	57	-0	-0	-0	-0	-0	-191	-329	-599

Notes: The Table presents results on the economic value generated by ATSM forecasts using the extended estimation procedure. We report results of the in-sample and out-of-sample analysis for 25 horizon (τ) and maturity ($T-\tau$) combinations using the latent and observable factor model in Panels A and B, respectively. The horizons are indicated in the rows, the maturities in the columns. In each Panel, we first present the excess returns of optimal bond portfolios using extended estimation forecasts. For an investment horizon τ, an investor with target volatility of $\sigma^* = 2\%$ and relative risk aversion $\rho = 3$ optimally allocates his wealth between the risk-free bond with maturity τ and the risky bond with maturity $T > \tau$. In the remaining columns, we report the economic value added by the extended compared to the standard estimation and compared to the EH-consistent constant risk premium forecasts using the performance measure of Goetzmann et al. (2007). All performance measures are annualized and reported in basis points. For the in-sample analysis, we estimate the models using monthly US data from January 1952 to December 2012. In the out-of-sample analysis, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t. We evaluate the economic value of forecasts that have their corresponding realization occurring during the financial crisis between July 2007 and December 2012.

Table IA.6: Directional Accuracy of Bond Excess Return Forecasts (Crisis Subsample 1)

Panel A: $A_0(3)$ Model with Latent Factors

	In-Sample					Out-of-Sample				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	0.51°	0.82°	1	1	$1.04^{\star}$	0.91	0.92°	1	1	0.96
3m	0.81°	0.92°	1	1	1	0.92°	1	1	1	0.94
12m	1°	1°	1	1	1	0.98°	0.98°	0.89°	0.85°	0.78°
24m	1.03	1.03	1°	0.98°	0.94°	0.94°	0.88°	0.77°	0.7°	0.66°
60m	0.92°	0.89°	0.77°	0.71°	$0.55^{\circ\circ}$	0.83°	0.74°	$0.58^{\circ\circ}$	$0.5^{\circ\circ}$	$0.26^{\circ\circ}$

Panel B: $A_0(3)$ Model with Observable Factors

	In-Sample					Out-of-Sample				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	0.51°	$0.37^{\circ\circ}$	$0.49^{\circ\circ}$	0.71°	0.83°	$0.45^{\circ\circ}$	$0.31^{\circ\circ}$	$0.43^{\circ\circ}$	0.73°	0.83°
3m	$0.42^{\circ\circ}$	$0.22^{\circ\circ}$	$0.41^{\circ\circ}$	0.69°	0.83°	$0.36^{\circ\circ}$	$0.19^{\circ\circ}$	$0.37^{\circ\circ}$	0.71°	0.83°
12m	$0.57^{\circ\circ}$	$0.57^{\circ\circ}$	$0.58^{\circ\circ}$	$0.62^{\circ\circ}$	0.7°	$0.57^{\circ\circ}$	$0.54^{\circ\circ}$	$0.55^{\circ\circ}$	0.62°	0.7°
24m	1°	0.95°	0.66°	0.59°	0.59°	0.98°	0.88°	0.59°	0.59°	0.59°
60m	1°	0.98°	0.62°	$0.52^{\circ\circ}$	$0.5^{\circ\circ}$	0.86°	0.71°	$0.42^{\circ\circ}$	$0.38^{\circ\circ}$	$0.3^{\circ\circ}$

Notes: The Table presents measures of the directional accuracy of extended estimation compared to EH-consistent constant risk premium forecasts based on hit ratios that measuring the ratio of correctly signed forecasts. Panels A and B contain results for the latent and observable factor model, respectively. Values reported are computed as the fraction of the ATSM model hit ratio relative to the EH hit ratio. Estimates with a * indicate that the model hit ratio exceeds the 95%-percentile of the bootstrapped constant risk premium hit ratio distribution. ** indicates that the 5%-percentile of the model distribution is higher than the 95%-percentile of the constant risk premium hit ratio distribution. $^{\circ}$ and $^{\circ\circ}$ indicate analogous results when the model hit ratio is lower than the constant risk premium hit ratio. We present in- and out-of-sample results for 25 horizon (τ) and maturity ($T-\tau$) combinations, respectively. The horizons are indicated in the rows, the maturities in the columns. For the in-sample analysis, we estimate the models using monthly US data from January 1952 to December 2012. In the out-of-sample analysis, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t. We compute hit ratios for the forecasts that have their corresponding realizations occuring during the financial crisis between July 2007 and December 2012.

Table IA.7: Statistical Accuracy of Bond Excess Return Forecasts (Crisis Subsample 2)

Panel A: $A_0(3)$ Model with Latent Factors

	In-Sample										Out-of-Sample									
	Extended vs. Standard Estimation					Model vs. EH Forecast					Extended vs. Standard Estimation					Model vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	−0.02	0.33	0.31	0.3	0.2*	0.3*	0.3*	0.19	0.23	0.18	−0.02	0.09	0.14	0.14*	0.1*	0.3*	0.05	−0.06	−0.05°	−0.02
3m	0.53	0.61	0.44	0.51	0.43*	0.53*	0.47*	0.33	0.44	0.43*	0.15	0.39*	0.27*	0.26*	0.22*	0.12	0.06	−0.16°	−0.12°	−0.06°
12m	0.58	0.43	0.39	0.84	0.7*	0.38	0.22	0.32	0.84	0.78*	0.57*	0.62*	0.71*	0.75*	0.56*	−0.32	−0.2°	0.04	0.29	0.04
24m	0.82	0.83	0.95	0.91*	0.23	0.76	0.82	0.96*	0.94*	0.62*	0.74*	0.77*	0.81*	0.75*	0.5*	−0.29°	0.08	0.41*	0.37*	−0.01
60m	0.74*	0.72*	0.65*	0.52*	0.2	0.67*	0.69*	0.63*	0.49*	0.15*	0.57*	0.54*	0.43*	0.28*	0*	−0.06	−0.01	−0.12	−0.31	−0.68°

Panel B: $A_0(3)$ Model with Observable Factors

	In-Sample										Out-of-Sample									
	Extended vs. Standard Estimation					Model vs. EH Forecast					Extended vs. Standard Estimation					Model vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	−0.08°	−0.06°	−0.01°	0	0.01*	0.14	−0.27	−0.34	−0.14°	0.02	−0.08°	−0.04°	0.04	0.04*	0.02*	0.08	−0.36	−0.39°	−0.15°	0.02
3m	−0.11°	−0.07°	−0.03°	0	0.02*	−0.23	−0.68	−0.4°	−0.2°	0.04	−0.08°	−0.03°	0.04	0.06*	0.06*	−0.34	−0.81	−0.44°	−0.2°	0.05
12m	−0.12°	−0.11°	−0.08°	−0.03°	0.06*	−0.97	−0.92	−0.68°	−0.31	0.26*	0	0.02	0.08	0.13*	0.21*	−1.02	−0.95	−0.64°	−0.26	0.27*
24m	−0.1°	−0.1°	−0.08°	−0.05°	0.04*	−1.23°	−0.87°	−0.4	−0.12	0.29*	0.08	0.08	0.1*	0.14*	0.22*	−1.2°	−0.84°	−0.38	−0.12	0.25*
60m	−0.01°	−0.02°	−0.03°	−0.02°	−0.01°	−1.02°	−0.86°	−0.73°	−0.64°	−0.51°	0.16*	0.16*	0.14*	0.14*	0.14*	−1.25°	−1.07°	−0.94°	−0.83°	−0.71°

Notes: The Table presents $R2$ statistics measuring the predictive ability of the ATSM estimated with the extended procedure relative to the standard estimation and EH-consistent constant risk premium benchmarks. The $R2$ statistic is defined as $R2 \equiv 1 - MSE^m/MSE^b$ where MSE^m and MSE^b denote the mean squared errors of the model and benchmark forecasts, respectively. We report in-sample and out-of-sample results for 25 horizon (τ) and maturity ($T-\tau$) combinations for the latent and observable factor model in Panels A and B, respectively. The horizons are indicated in the rows, the maturities in the columns. * ($^\circ$) identifies $R2$s that are significantly positive (negative) as judged by confidence intervals obtained as the 5%- and 95%-percentiles from a block bootstrap procedure. For the in-sample analysis, we estimate the models using monthly US data from January 1952 to December 2012. In the out-of-sample analysis, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t. We measure the statistical accuracy of forecasts made and corresponding realizations occurring during the financial crisis between July 2007 and December 2012.

Table IA.8: Economic Value of Bond Excess Return Forecasts (Crisis Subsample 2)

Panel A: $A_0(3)$ Model with Latent Factors

	In-Sample															Out-of-Sample													
	Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast					Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	9	63	156	254	508	2	41	135	207	350	-7	-3	-0	-0	22	14	64	152	247	376	1	17	111	295	499	-1	-2	-0	-0	-105
3m	21	87	140	233	454	5	44	88	137	328	-1	-2	-0	-0	-6	21	90	136	227	392	9	70	164	303	670	-1	-0	-0	-0	-65
12m	18	55	128	224	440	-0	4	15	74	126	-0	-0	-0	-0	-0	18	55	125	221	387	19	58	186	394	800	-0	-0	-0	-0	-53
24m	17	54	139	230	419	-0	-0	-0	12	60	-0	-0	-0	-0	-0	17	54	137	214	366	20	61	212	392	748	-0	-0	-0	-15	-60
60m	15	46	141	213	359	-0	-0	-0	-0	137	-0	-0	-0	-0	-0	15	46	106	39	-360	5	15	161	175	-0	-0	-0	-30	-156	-602

Panel B: $A_0(3)$ Model with Observable Factors

	In-Sample															Out-of-Sample													
	Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast					Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	3	7	-50	38	181	2	-0	-13	-17	-4	-13	-59	-208	-222	-314	-3	-20	-93	46	179	-0	-0	-0	149	68	-18	-86	-247	-208	-309
3m	10	21	-9	48	163	1	-6	-5	9	8	-12	-68	-150	-192	-310	5	10	-20	50	161	-3	10	30	66	57	-17	-80	-158	-184	-307
12m	12	34	65	104	296	-3	-8	-6	-3	9	-7	-22	-63	-123	-151	12	34	54	103	296	1	3	14	26	34	-7	-22	-71	-120	-150
24m	17	54	85	134	296	-0	-0	-11	-0	-0	-0	-0	-52	-98	-129	17	54	75	133	298	3	14	10	8	11	-0	-0	-61	-97	-130
60m	15	46	73	-132	-359	-0	-0	-59	-108	0	-0	-0	-59	-290	-600	15	46	-92	-187	-360	13	57	-0	-0	-0	-0	-0	-191	-331	-602

Notes: The Table presents results on the economic value generated by ATSM forecasts using the extended estimation procedure. We report results of the in-sample and out-of-sample analysis for 25 horizon (τ) and maturity ($T-\tau$) combinations using the latent and observable factor model in Panels A and B, respectively. The horizons are indicated in the rows, the maturities in the columns. In each Panel, we first present the excess returns of optimal bond portfolios using extended estimation forecasts. For an investment horizon τ, an investor with target volatility of $\sigma^* = 2\%$ and relative risk aversion $\rho = 3$ optimally allocates his wealth between the risk-free bond with maturity τ and the risky bond with maturity $T > \tau$. In the remaining columns, we report the economic value added by the extended compared to the standard estimation and compared to the EH-consistent constant risk premium forecasts using the performance measure of Goetzmann et al. (2007). All performance measures are annualized and reported in basis points. For the in-sample analysis, we estimate the models using monthly US data from January 1952 to December 2012. In the out-of-sample analysis, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t. We measure the economic value of forecasts made and corresponding realizations occurring during the financial crisis between July 2007 and December 2012.

Table IA.9: Directional Accuracy of Bond Excess Return Forecasts (Crisis Subsample 2)

Panel A: $A_0(3)$ Model with Latent Factors

	In-Sample					Out-of-Sample				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	0.51°	0.82°	1	1	$1.04^{\star}$	0.91	0.92	1	1	0.98
3m	0.8°	0.92°	1	1	1	0.92°	1	1	1	0.96
12m	1	1	1	1	1	1	1	1	1	0.96
24m	1	1	1	1	1	1	1	1	0.98°	0.95°
60m	1	1	1	1	1	1	1	0.83°	0.67°	$0^{\circ\circ}$

Panel B: $A_0(3)$ Model with Observable Factors

	In-Sample					Out-of-Sample				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
τ =1m	0.51°	$0.37^{\circ\circ}$	$0.48^{\circ\circ}$	0.72°	0.84°	$0.45^{\circ\circ}$	$0.31^{\circ\circ}$	$0.42^{\circ\circ}$	0.74°	0.84°
3m	$0.39^{\circ\circ}$	$0.18^{\circ\circ}$	$0.37^{\circ\circ}$	0.7°	0.84°	$0.33^{\circ\circ}$	$0.15^{\circ\circ}$	$0.33^{\circ\circ}$	0.72°	0.84
12m	$0.48^{\circ\circ}$	$0.52^{\circ\circ}$	$0.67^{\circ\circ}$	$0.75^{\circ\circ}$	0.87°	$0.48^{\circ\circ}$	$0.52^{\circ\circ}$	$0.67^{\circ\circ}$	0.75°	0.87°
24m	1	1	0.88°	0.86°	0.86°	1	1	0.86°	0.86°	0.86°
60m	1	1	0.67°	$0.17^{\circ\circ}$	$0^{\circ\circ}$	1	1	$0^{\circ\circ}$	$0^{\circ\circ}$	$0^{\circ\circ}$

Notes: The Table presents measures of the directional accuracy of extended estimation compared to EH-consistent constant risk premium forecasts based on hit ratios that measuring the ratio of correctly signed forecasts. Panels A and B contain results for the latent and observable factor model, respectively. Values reported are computed as the fraction of the ATSM model hit ratio relative to the EH hit ratio. Estimates with a * indicate that the model hit ratio exceeds the 95%-percentile of the bootstrapped constant risk premium hit ratio distribution. ** indicates that the 5%-percentile of the model distribution is higher than the 95%-percentile of the constant risk premium hit ratio distribution. $^{\circ}$ and $^{\circ\circ}$ indicate analogous results when the model hit ratio is lower than the constant risk premium hit ratio. We present in- and out-of-sample results for 25 horizon (τ) and maturity ($T-\tau$) combinations, respectively. The horizons are indicated in the rows, the maturities in the columns. For the in-sample analysis, we estimate the models using monthly US data from January 1952 to December 2012. In the out-of-sample analysis, we first estimate the model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t. We compute hit ratios of forecasts made and corresponding realizations occurring during the financial crisis between July 2007 and December 2012.

Table IA.10: Yield Pricing Errors: International Evidence

Yield Maturity		All T	$1\text{m} \leq T < 12\text{m}$	$12\text{m} \leq T < 60\text{m}$	$60\text{m} \leq T \leq 120\text{m}$
Number of Yields		24	7	10	7
Switzerland					
RMSE	standard estimation	6.80	7.54	6.77	6.03
	extended estimation	11.92	12.49	11.21	12.30
Sd	standard estimation	6.80	7.54	6.77	6.03
	extended estimation	11.92	12.49	11.21	12.29
Germany					
RMSE	standard estimation	7.88	8.64	7.07	8.17
	extended estimation	19.74	33.34	10.02	8.96
Sd	standard estimation	7.88	8.65	7.06	8.11
	extended estimation	19.74	33.34	10.01	8.96
United Kingdom					
RMSE	standard estimation	11.02	15.95	8.94	6.89
	extended estimation	23.48	39.72	12.05	10.27
Sd	standard estimation	11.01	15.94	8.93	6.88
	extended estimation	23.46	39.61	12.03	10.22
Japan					
RMSE	standard estimation	5.84	5.63	5.86	6.02
	extended estimation	10.46	15.94	7.12	6.98
Sd	standard estimation	5.84	5.63	5.85	6.03
	extended estimation	10.45	15.90	7.12	6.98
United States					
RMSE	standard estimation	10.16	13.18	9.28	7.53
	extended estimation	15.85	25.36	10.65	7.48
Sd	standard estimation	10.15	13.17	9.28	7.41
	extended estimation	15.85	25.31	10.62	7.48

Notes: The Table summarizes root mean squared yield pricing errors and standard deviations of yield pricing errors for the standard estimation (the estimation procedure only fitting yields) and the extended estimation (the estimation procedure fitting yields and matching model risk premia to bond excess returns observed in the past) of the $A_0(3)$ model. We estimate the models using monthly data from September 1989 for Japan, January 1988 for Switzerland, and April 1987 for Germany, UK, and US. The sample period ends in March 2011 for all countries.

Table IA.11: Time-Varying Risk Premium Regressions: International Evidence

Panel A: Standard Estimation Procedure

		$T-\tau$	1m	3m	12m	24m	60m
Switzerland							
τ	1m	b	0.44	0.74	0.51	1.33**	0.85**
		R^2	0.01	0.02	0.01	0.07	0.04
	3m	b	0.54	0.52	0.54	0.85	0.64*
		R^2	0.03	0.03	0.03	0.06	0.06
	12m	b	−0.05	0.28	0.28	0.46	0.45
		R^2	0.00	0.01	0.01	0.04	0.06
	24m	b	0.86	0.35	0.10	0.08	0.14
		R^2	0.07	0.02	0.00	0.00	0.01
	60m	b	−0.04	−0.33	−0.41	−0.30	−0.20
		R^2	0.00	0.07	0.14	0.09	0.06
Germany							
τ	1m	b	0.37	0.50	−0.27	0.12	0.51
		R^2	0.01	0.02	0.00	0.00	0.01
	3m	b	0.48	0.44	0.15	0.38	0.61
		R^2	0.03	0.02	0.00	0.01	0.02
	12m	b	−0.65	−0.06	0.20	0.40	0.61
		R^2	0.06	0.00	0.01	0.02	0.06
	24m	b	0.10	0.05	0.07	0.19	0.41
		R^2	0.00	0.00	0.00	0.01	0.06
	60m	b	0.44	0.01	−0.11	0.01	0.18
		R^2	0.04	0.00	0.01	0.00	0.05
United Kingdom							
τ	1m	b	1.16*	0.90*	1.30**	1.56**	1.26*
		R^2	0.02	0.02	0.04	0.06	0.03
	3m	b	0.73	0.80	0.78	0.87	0.76
		R^2	0.03	0.04	0.04	0.05	0.03
	12m	b	1.00*	0.64	0.31	0.26	0.37
		R^2	0.08	0.05	0.01	0.01	0.02
	24m	b	0.77	0.32	0.00	0.13	0.30
		R^2	0.05	0.01	0.00	0.00	0.02
	60m	b	0.14	0.03	0.04	0.25	0.48*
		R^2	0.00	0.00	0.00	0.04	0.20
Japan							
τ	1m	b	1.18**	1.18*	0.39	0.33	0.96*
		R^2	0.09	0.06	0.01	0.01	0.05
	3m	b	1.26	0.86	0.76	0.81	1.05*
		R^2	0.12	0.05	0.07	0.09	0.13
	12m	b	0.11	0.50	0.81	0.85	0.83
		R^2	0.00	0.06	0.22	0.29	0.33
	24m	b	−0.03	0.41	0.69	0.75	0.76**
		R^2	0.00	0.13	0.30	0.38	0.48
	60m	b	0.89**	0.65	0.53	0.62	0.62
		R^2	0.16	0.27	0.18	0.25	0.28
United States							
τ	1m	b	0.15	0.36	1.13	−0.40	0.77
		R^2	0.00	0.00	0.02	0.00	0.01
	3m	b	0.33	0.31	0.84	0.13	0.55
		R^2	0.01	0.01	0.02	0.00	0.01
	12m	b	0.88	0.85	1.42	0.95	0.72
		R^2	0.03	0.04	0.10	0.05	0.05
	24m	b	−1.97	0.11	1.19	0.92	0.70
		R^2	0.12	0.00	0.12	0.09	0.10
	60m	b	0.16	0.43	0.40	0.36	0.33
		R^2	0.00	0.05	0.06	0.06	0.11

Panel B: Extended Estimation Procedure

		$T-\tau$	1m	3m	12m	24m	60m
Switzerland							
τ	1m	b	0.98**	0.90**	0.96**	1.21**	1.05**
		R^2	0.25	0.26	0.29	0.31	0.23
	3m	b	0.85**	0.87**	1.00**	1.16**	1.07**
		R^2	0.54	0.57	0.61	0.60	0.52
	12m	b	0.98**	1.00**	1.02**	1.11**	1.04**
		R^2	0.73	0.88	0.84	0.87	0.82
	24m	b	1.14**	1.02**	0.94**	0.94**	0.89**
		R^2	0.42	0.65	0.62	0.61	0.62
	60m	b	0.59	0.70**	0.71**	0.61*	0.59**
		R^2	0.06	0.27	0.35	0.29	0.37
Germany							
τ	1m	b	0.91**	0.79**	0.87**	1.16**	1.14**
		R^2	0.22	0.20	0.22	0.27	0.28
	3m	b	0.80**	0.83**	0.99**	1.15**	1.04**
		R^2	0.33	0.40	0.51	0.55	0.49
	12m	b	1.02**	1.07**	1.14**	1.20**	1.02**
		R^2	0.60	0.84	0.83	0.84	0.77
	24m	b	1.28**	1.20**	1.11**	1.06**	0.90**
		R^2	0.30	0.56	0.58	0.56	0.53
	60m	b	0.96	0.83*	0.60**	0.52**	0.46*
		R^2	0.07	0.18	0.17	0.16	0.18
United Kingdom							
τ	1m	b	0.93**	0.86**	1.08**	1.19**	0.91**
		R^2	0.10	0.20	0.23	0.28	0.18
	3m	b	0.91**	0.92**	1.01**	1.04**	0.90**
		R^2	0.44	0.46	0.47	0.51	0.38
	12m	b	1.25**	1.13**	1.12**	1.10**	0.97**
		R^2	0.62	0.85	0.86	0.86	0.78
	24m	b	1.48**	1.23**	1.05**	0.98**	0.84**
		R^2	0.53	0.66	0.57	0.56	0.53
	60m	b	0.24	0.44	0.47	0.48	0.46*
		R^2	0.01	0.12	0.17	0.23	0.37
Japan							
τ	1m	b	0.60**	0.77**	1.04**	1.13**	1.34**
		R^2	0.25	0.41	0.50	0.37	0.34
	3m	b	0.72**	0.83**	1.07**	1.22**	1.39**
		R^2	0.54	0.69	0.82	0.75	0.64
	12m	b	1.24**	1.11**	1.14**	1.14**	1.08**
		R^2	0.61	0.78	0.81	0.80	0.70
	24m	b	0.08	0.71**	1.07**	1.07**	0.97**
		R^2	0.00	0.51	0.87	0.87	0.74
	60m	b	0.26	0.88**	1.11**	1.13**	1.08**
		R^2	0.02	0.60	0.93	0.92	0.82
United States							
τ	1m	b	0.73**	0.74**	1.00**	1.11**	0.98**
		R^2	0.16	0.21	0.29	0.26	0.21
	3m	b	0.83**	0.86**	1.02**	1.09**	0.97**
		R^2	0.46	0.52	0.59	0.56	0.45
	12m	b	1.38**	1.22**	1.20**	1.16**	0.96**
		R^2	0.65	0.85	0.86	0.86	0.74
	24m	b	1.20**	1.27**	1.25**	1.14**	0.91**
		R^2	0.21	0.55	0.61	0.60	0.58
	60m	b	0.12	0.36	0.31	0.27	0.30
		R^2	0.00	0.03	0.03	0.03	0.06

Notes: The Table presents results for regressing realized bond excess returns on risk premia implied by the $A_0(3)$ model using the standard estimation procedure that only fits yields (Panel A) and using the extended estimation procedure that fits yields and matches model risk premia to bond excess returns observed in the past (Panel B). For each country, we report estimates for 25 horizon (τ) and maturity ($T-\tau$) combinations. The horizons are indicated in the rows, the maturities in the columns. b is the estimate of the slope coefficient. ** and * indicates that the estimate differs from zero at the 99% or 95% level, respectively. Significance is assessed using standard errors with HAC adjustment based on Newey and West (1987) where the optimal truncation lag is chosen following Andrews (1991) using a quadratic spectral kernel (standard errors not reported for space reasons). R^2 denotes the in-sample coefficient of determination. We estimate the models using monthly data from September 1989 for Japan, January 1988 for Switzerland, and April 1987 for Germany, UK, and US. The sample period ends in March 2011 for all countries.

Table IA.12: Statistical Accuracy of Bond Excess Return Forecasts: International Evidence

		Panel A: In-Sample										Panel B: Out-of-Sample									
		Extended vs. Standard Estimation					Extended Estimation vs. EH					Extended vs. Standard Estimation					Extended Estimation vs. EH				
	$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
Switzerland																					
τ	1m	$0.23^{\star}$	$0.25^{\star}$	$0.30^{\star}$	$0.21^{\star}$	$0.18^{\star}$	$0.23^{\star}$	$0.26^{\star}$	$0.29^{\star}$	$0.25^{\star}$	$0.22^{\star}$	-0.04	0.02	0	-0.02	-0.22°	-0.38°	-0.24°	-0.10°	-0.26°	-0.31°
	3m	$0.53^{\star}$	$0.57^{\star}$	$0.62^{\star}$	$0.55^{\star}$	$0.50^{\star}$	$0.52^{\star}$	$0.55^{\star}$	$0.61^{\star}$	$0.57^{\star}$	$0.51^{\star}$	$0.16^{\star}$	0.1	0.01	-0.05	-0.24°	-0.29°	-0.24	-0.25	-0.34°	-0.41°
	12m	$0.77^{\star}$	$0.89^{\star}$	$0.86^{\star}$	$0.87^{\star}$	$0.84^{\star}$	$0.73^{\star}$	$0.88^{\star}$	$0.84^{\star}$	$0.86^{\star}$	$0.82^{\star}$	0.35	$0.46^{\star}$	$0.42^{\star}$	0.35	0.22	0.02	0.05	0.07	0.08	-0.01
	24m	0.31	$0.65^{\star}$	$0.69^{\star}$	$0.70^{\star}$	$0.73^{\star}$	0.31	$0.63^{\star}$	$0.62^{\star}$	$0.61^{\star}$	$0.61^{\star}$	$0.54^{\star}$	$0.67^{\star}$	$0.60^{\star}$	$0.57^{\star}$	$0.54^{\star}$	0.02	$0.15^{\star}$	$0.20^{\star}$	0.13	0.02
	60m	0.22	$0.63^{\star}$	$0.73^{\star}$	$0.71^{\star}$	$0.77^{\star}$	0.03	0.21	0.29	0.17	0.18	$0.67^{\star}$	$0.84^{\star}$	$0.85^{\star}$	$0.84^{\star}$	$0.87^{\star}$	-0.01	0.15	$0.31^{\star}$	0.26	0.36
Germany																					
τ	1m	0.23	0.18	$0.25^{\star}$	$0.26^{\star}$	$0.27^{\star}$	0.19	0.18	$0.21^{\star}$	$0.20^{\star}$	$0.24^{\star}$	0.13	$0.18^{\star}$	-0.11	$0.19^{\star}$	0.11	0.04	0.26	0.05	0.14	0.07
	3m	0.32	0.38	$0.54^{\star}$	$0.54^{\star}$	$0.48^{\star}$	0.31	0.37	$0.51^{\star}$	$0.52^{\star}$	$0.48^{\star}$	$0.34^{\star}$	0.19	0	0.19	0.06	0.31	0.26	0.04	0.15	0.04
	12m	$0.69^{\star}$	$0.85^{\star}$	$0.83^{\star}$	$0.82^{\star}$	$0.76^{\star}$	$0.59^{\star}$	$0.82^{\star}$	$0.81^{\star}$	$0.81^{\star}$	$0.77^{\star}$	0.18	0.28	0.06	0.12	0.11	0.15	0.2	-0.05	0.02	0.03
	24m	$0.33^{\star}$	$0.63^{\star}$	$0.64^{\star}$	$0.61^{\star}$	$0.55^{\star}$	0.18	$0.54^{\star}$	$0.56^{\star}$	$0.55^{\star}$	$0.52^{\star}$	$0.42^{\star}$	$0.53^{\star}$	0.21	0.23	0.26	0.15	$0.33^{\star}$	-0.09	-0.08	-0.03
	60m	0.13	$0.51^{\star}$	$0.63^{\star}$	$0.60^{\star}$	$0.55^{\star}$	-0.08	0.15	0.09	0.02	-0.06	$0.54^{\star}$	0.37	-0.41	-0.15	0.36	-0.12	-0.72°	-1.25°	-1.07°	-0.71°
United Kingdom																					
τ	1m	$0.10^{\star}$	$0.19^{\star}$	$0.21^{\star}$	$0.25^{\star}$	$0.15^{\star}$	$0.09^{\star}$	$0.20^{\star}$	$0.22^{\star}$	$0.27^{\star}$	$0.17^{\star}$	0.04	$0.30^{\star}$	0.25	$0.30^{\star}$	0.18	0.07	0.29	0.26	0.27	0.08
	3m	$0.44^{\star}$	$0.44^{\star}$	$0.47^{\star}$	$0.50^{\star}$	$0.36^{\star}$	$0.43^{\star}$	$0.45^{\star}$	$0.47^{\star}$	$0.51^{\star}$	$0.37^{\star}$	$0.44^{\star}$	0.34	0.23	0.21	0.01	0.41	0.32	0.17	0.11	-0.23
	12m	$0.61^{\star}$	$0.84^{\star}$	$0.87^{\star}$	$0.87^{\star}$	$0.80^{\star}$	$0.59^{\star}$	$0.83^{\star}$	$0.85^{\star}$	$0.85^{\star}$	$0.78^{\star}$	$0.41^{\star}$	$0.60^{\star}$	$0.52^{\star}$	$0.46^{\star}$	0.27	$0.40^{\star}$	$0.52^{\star}$	$0.33^{\star}$	$0.18^{\star}$	-0.25
	24m	$0.54^{\star}$	$0.72^{\star}$	$0.70^{\star}$	$0.68^{\star}$	$0.60^{\star}$	$0.47^{\star}$	$0.63^{\star}$	$0.57^{\star}$	$0.56^{\star}$	$0.51^{\star}$	$0.43^{\star}$	$0.62^{\star}$	$0.54^{\star}$	$0.54^{\star}$	$0.52^{\star}$	$0.29^{\star}$	$0.39^{\star}$	0.2	0.18	0.07
	60m	$0.55^{\star}$	$0.73^{\star}$	$0.71^{\star}$	$0.69^{\star}$	$0.61^{\star}$	-0.15	-0.07	-0.05	-0.05	-0.14	0.17	$0.53^{\star}$	$0.55^{\star}$	$0.59^{\star}$	$0.62^{\star}$	-0.31	-0.25	-0.14	-0.07	-0.13
Japan																					
τ	1m	0.07	$0.34^{\star}$	$0.50^{\star}$	$0.33^{\star}$	$0.27^{\star}$	0.11	$0.37^{\star}$	$0.49^{\star}$	0.23	$0.30^{\star}$	$0.25^{\star}$	$0.40^{\star}$	0.25	$0.50^{\star}$	$0.29^{\star}$	-0.37°	-0.09	0.02	-0.71°	-0.02
	3m	$0.40^{\star}$	$0.64^{\star}$	$0.80^{\star}$	$0.68^{\star}$	$0.52^{\star}$	$0.45^{\star}$	$0.66^{\star}$	$0.81^{\star}$	$0.70^{\star}$	$0.58^{\star}$	$0.41^{\star}$	$0.49^{\star}$	$0.39^{\star}$	$0.64^{\star}$	$0.36^{\star}$	-0.24	-0.04	0.04	0.19	0.16
	12m	$0.63^{\star}$	$0.77^{\star}$	$0.72^{\star}$	$0.70^{\star}$	$0.56^{\star}$	$0.59^{\star}$	$0.77^{\star}$	$0.77^{\star}$	$0.78^{\star}$	$0.70^{\star}$	0.23	$0.54^{\star}$	-0.91°	-0.17	$0.64^{\star}$	0.11	$0.46^{\star}$	$0.57^{\star}$	$0.66^{\star}$	$0.70^{\star}$
	24m	-0.03	0.4	$0.81^{\star}$	$0.80^{\star}$	$0.54^{\star}$	-1.16°	0.23	$0.84^{\star}$	$0.86^{\star}$	$0.73^{\star}$	$0.42^{\star}$	$0.76^{\star}$	-1.22°	-0.36	$0.65^{\star}$	-2.44°	$0.57^{\star}$	$0.71^{\star}$	$0.77^{\star}$	$0.82^{\star}$
	60m	-0.61°	$0.62^{\star}$	$0.93^{\star}$	$0.93^{\star}$	$0.85^{\star}$	-0.41°	$0.57^{\star}$	$0.86^{\star}$	$0.89^{\star}$	$0.80^{\star}$	$0.39^{\star}$	$0.63^{\star}$	-3.97°	-2.46°	0.24	-0.5	$0.80^{\star}$	$0.70^{\star}$	$0.78^{\star}$	$0.88^{\star}$
United States																					
τ	1m	0.15	$0.19^{\star}$	$0.27^{\star}$	$0.25^{\star}$	$0.20^{\star}$	0.13	$0.18^{\star}$	$0.29^{\star}$	$0.21^{\star}$	$0.21^{\star}$	0.21	0.18	$0.28^{\star}$	$0.27^{\star}$	0.05	0.11	0.1	$0.25^{\star}$	$0.16^{\star}$	0.00
	3m	$0.46^{\star}$	$0.51^{\star}$	$0.58^{\star}$	$0.55^{\star}$	$0.44^{\star}$	$0.44^{\star}$	$0.50^{\star}$	$0.59^{\star}$	$0.54^{\star}$	$0.45^{\star}$	0.26	0.17	0.16	$0.21^{\star}$	-0.05	0.16	0.1	0.12	0.15	-0.07
	12m	$0.58^{\star}$	$0.82^{\star}$	$0.82^{\star}$	$0.83^{\star}$	$0.73^{\star}$	$0.59^{\star}$	$0.82^{\star}$	$0.83^{\star}$	$0.84^{\star}$	$0.74^{\star}$	$0.48^{\star}$	$0.42^{\star}$	0.15	0.24	0.17	$0.46^{\star}$	$0.37^{\star}$	0.11	0.19	0.15
	24m	$0.25^{\star}$	$0.55^{\star}$	$0.56^{\star}$	$0.58^{\star}$	$0.55^{\star}$	0.1	$0.52^{\star}$	$0.58^{\star}$	$0.58^{\star}$	$0.57^{\star}$	$0.40^{\star}$	$0.57^{\star}$	$0.40^{\star}$	$0.42^{\star}$	$0.34^{\star}$	$0.19^{\star}$	$0.40^{\star}$	$0.19^{\star}$	$0.20^{\star}$	0.06
	60m	-0.07	0.06	0.11	0.05	0.05	-0.14	-0.08	-0.16	-0.2	-0.27	0.2	$0.43^{\star}$	$0.44^{\star}$	$0.39^{\star}$	0.11	-1.08°	-1.49°	-1.35°	-2.13°	-6.09°

Notes: The Table presents $R2$ statistics measuring the predictive ability of the ATSM estimated with the extended procedure relative to the standard estimation and EH-consistent constant risk premium benchmarks. The $R2$ statistic is defined as $R2 \equiv 1 - MSE^m/MSE^b$ where MSE^m and MSE^b denote the mean squared errors of the model and benchmark forecasts, respectively. Panels A and B report in-sample and out-of-sample results for 25 horizon (τ) and maturity ($T-\tau$) combinations, respectively. The horizons are indicated in the rows, the maturities in the columns. * ($^{\circ}$) identifies $R2$s that are significantly positive (negative) as judged by confidence intervals obtained as the 5%- and 95%-percentiles from a block bootstrap procedure. For the in-sample analysis, we estimate the models using monthly data from September 1989 for Japan, January 1988 for Switzerland, and April 1987 for Germany, UK, and US. The sample period ends in March 2011 for all countries. In the out-of-sample analysis, we estimate for each country data set the first model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t.

Table IA.13: Economic Value of Bond Excess Return Forecasts: International Evidence

	Panel A: In-Sample															Panel B: Out-of-Sample													
	Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast					Portfolio Excess Returns					Extended vs. Standard Estimation					Extended Estimation vs. EH Forecast				
$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
Switzerland																														
τ =1m	68	122	245	489	633	21	48	148	81	116	23	69	157	153	267	-34	-17	-16	91	-64	-10	4	33	-68	-306	-101	-98	-109	-429	-626
3m	61	104	229	367	526	25	47	109	128	220	34	71	139	181	282	11	23	26	2	-26	25	41	36	-65	-167	-30	-17	-90	-259	-331
12m	39	80	158	225	371	22	40	84	111	184	20	40	79	114	177	21	45	78	114	132	31	52	74	88	42	-1	-4	-1	1	-51
24m	37	69	124	175	290	2	12	32	44	77	5	16	35	47	72	49	67	83	117	191	50	68	82	113	159	0	0	0	0	4
60m	27	82	185	272	466	0	0	0	0	0	0	0	0	0	0	28	73	156	235	405	14	44	88	118	171	0	0	0	0	0
Germany																														
τ = 1m	57	95	193	441	699	25	37	166	185	333	-1	27	84	53	200	29	111	140	560	663	32	75	27	146	177	-76	-2	72	-25	-7
3m	38	73	193	331	516	13	35	144	187	255	8	36	90	118	228	55	76	137	299	381	51	50	59	91	84	0	22	22	14	44
12m	34	76	164	233	384	35	61	123	163	200	14	32	73	110	169	18	52	79	118	197	14	25	27	37	55	2	-3	-8	-5	-9
24m	32	62	112	164	282	9	30	72	95	118	1	10	30	43	72	52	71	80	113	185	12	16	23	31	53	-1	-2	-2	-2	-2
60m	29	69	162	252	419	1	5	8	10	26	0	0	0	0	-4	35	68	139	219	375	0	0	0	0	0	0	0	0	0	0
United Kingdom																														
τ = 1m	44	81	197	363	508	28	43	93	199	231	20	48	88	161	226	31	83	225	358	410	-9	22	127	190	189	-14	48	74	101	70
3m	40	72	177	298	442	28	44	139	228	301	24	54	105	165	239	44	69	153	222	262	14	27	103	135	224	31	67	42	60	50
12m	31	69	152	229	362	22	54	126	185	245	14	40	82	111	176	42	63	108	142	208	21	41	85	131	234	15	18	24	22	48
24m	26	51	108	154	245	17	38	93	106	135	9	18	30	37	64	20	38	70	93	113	11	24	51	99	170	2	3	-6	-10	-13
60m	27	65	127	184	262	25	59	104	127	107	0	0	1	-7	-55	8	25	46	36	17	-5	5	13	3	135	-10	-19	-45	-95	-180
Japan																														
τ = 1m	81	159	358	699	768	8	27	142	136	184	18	66	129	76	208	23	76	142	519	388	7	28	89	475	424	-76	-53	-28	-64	-42
3m	65	116	263	433	559	1	25	89	115	149	17	38	79	96	187	32	62	94	227	214	5	26	77	178	190	-35	-24	-11	-4	-3
12m	43	88	157	240	437	9	13	18	27	63	4	9	16	24	60	23	40	38	72	195	8	26	16	28	98	-1	-2	0	0	7
24m	73	136	160	241	429	0	0	0	0	6	0	0	0	0	2	66	97	55	96	235	32	41	16	25	74	0	0	0	0	0
60m	41	90	149	220	382	0	0	0	0	0	0	0	0	0	0	38	66	75	118	228	2	0	0	0	3	0	0	0	0	0
United States																														
τ = 1m	69	144	329	496	659	1	42	67	44	215	-2	38	67	44	206	76	130	307	629	569	20	50	99	113	195	-18	-2	57	-29	5
3m	65	119	276	385	513	11	37	80	98	190	12	37	80	98	190	78	122	229	362	408	36	60	89	120	231	8	17	36	19	62
12m	49	95	192	270	413	7	16	45	72	115	7	16	45	72	115	53	93	151	223	346	7	16	38	58	134	7	11	21	34	66
24m	38	77	133	189	308	-1	-1	2	2	2	-1	-1	2	2	2	76	100	131	192	282	7	11	23	41	74	0	0	2	6	-2
60m	36	82	167	250	441	0	0	0	0	0	0	0	0	0	0	43	83	164	238	254	5	12	26	51	10	0	0	0	-5	-181
	111	-19	-41	-70	-54	-35																								

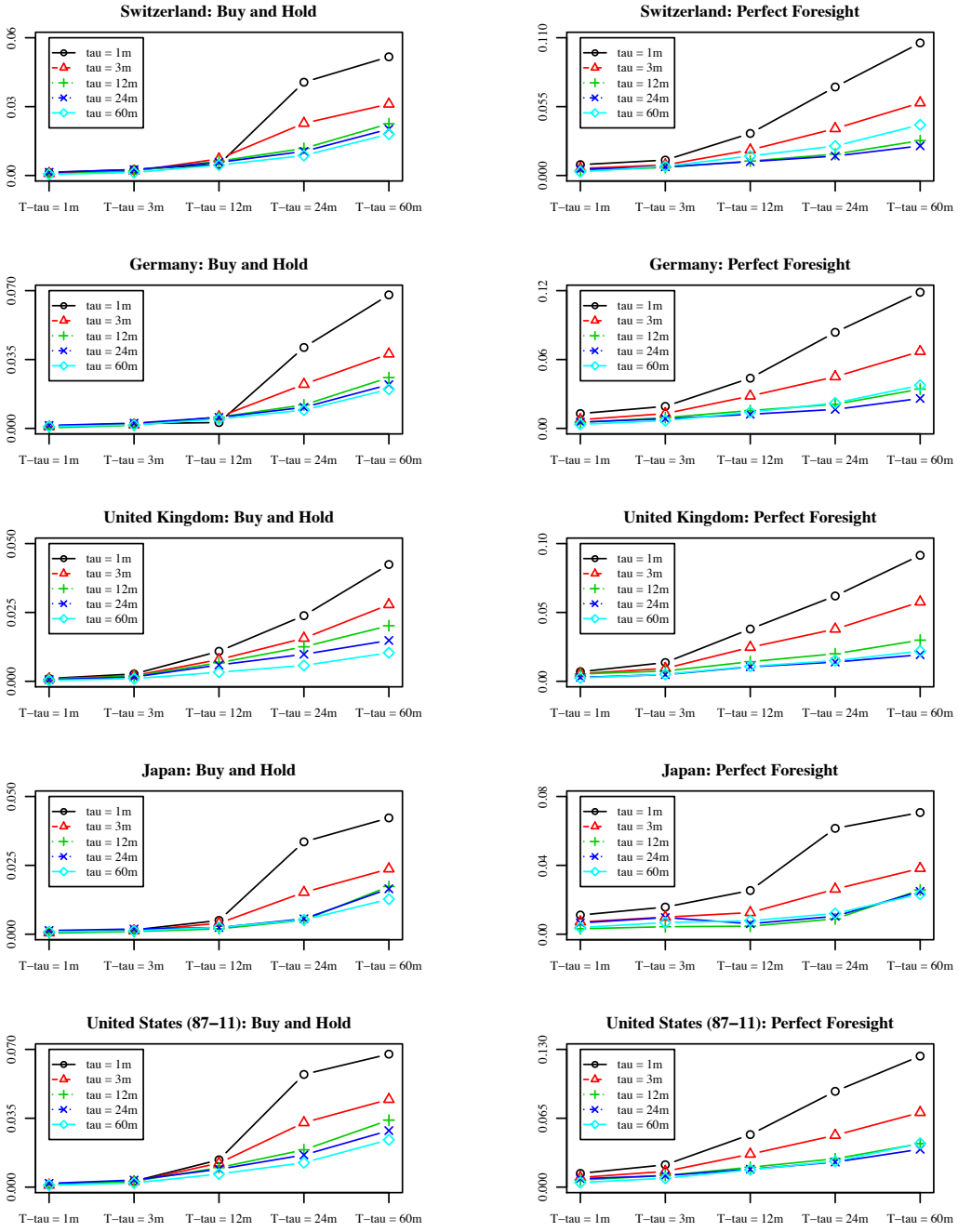
Notes: The Table presents results on the economic value generated by ATSM forecasts using the extended estimation procedure. We report results for the in-sample and out-of-sample analysis in Panels A and B for 25 horizon (τ) and maturity ($T-\tau$) combinations, respectively. The horizons are indicated in the rows, the maturities in the columns. In each Panel, we first present the excess returns of optimal bond portfolios using extended estimation forecasts. For an investment horizon τ, an investor with target volatility of $\sigma^* = 2\%$ and relative risk aversion $\rho = 3$ optimally allocates his wealth between the risk-free bond with maturity τ and the risky bond with maturity $T > \tau$. In the remaining columns, we report the economic value added by the extended compared to the standard estimation and compared to the EH-consistent constant risk premium forecasts using the performance measure of Goetzmann et al. (2007). All performance measures are annualized and reported in basis points. For the in-sample analysis, we estimate the models using monthly data from September 1989 for Japan, January 1988 for Switzerland, and April 1987 for Germany, UK, and US. The sample period ends in March 2011 for all countries. In the out-of-sample analysis, we estimate for each country data set the first model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t.

Table IA.14: Directional Accuracy of Bond Excess Return Forecasts: International Evidence

		In-Sample					Out-of-Sample				
	$T-\tau$	1m	3m	12m	24m	60m	1m	3m	12m	24m	60m
Switzerland											
τ	1m	0.99	1.07	1.15	1.03	$1.12^{\star}$	$0.45^{\circ\circ}$	$0.66^{\circ\circ}$	1.04	$0.72^{\circ\circ}$	$0.63^{\circ\circ}$
	3m	1.12	$1.18^{\star}$	$1.27^{\star\star}$	$1.25^{\star\star}$	$1.17^{\star}$	$0.63^{\circ\circ}$	0.84	0.81°	0.72°	$0.71^{\circ\circ}$
	12m	$1.31^{\star}$	$1.33^{\star\star}$	$1.30^{\star\star}$	$1.30^{\star\star}$	$1.32^{\star\star}$	0.98	1.00	1.00	1.03	0.93
	24m	1.04	1.09	1.17	1.17	1.16	1.00	1.00	1.00	1.00	1.01
	60m	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Germany											
τ	1m	0.91	1.02	1.13	1.01	1.08	$0.69^{\circ\circ}$	1.00	1.13	1.01	1.03
	3m	1.00	$1.18^{\star}$	$1.24^{\star}$	$1.21^{\star}$	$1.20^{\star}$	0.99	1.18	1.07	1.06	1.05
	12m	$1.23^{\star}$	$1.32^{\star\star}$	$1.31^{\star\star}$	$1.36^{\star\star}$	$1.30^{\star\star}$	1.05	0.97	0.96	0.99	0.99
	24m	1.03	1.12	1.14	1.16	$1.20^{\star}$	0.98	0.98	1.00	0.99	0.99
	60m	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
United Kingdom											
τ	1m	0.97	$1.14^{\star}$	$1.20^{\star}$	$1.14^{\star}$	$1.14^{\star}$	1.03	$1.41^{\star\star}$	$1.21^{\star}$	$1.21^{\star}$	1.08
	3m	$1.32^{\star\star}$	$1.45^{\star\star}$	$1.36^{\star\star}$	$1.28^{\star\star}$	$1.20^{\star}$	$1.33^{\star\star}$	1.15	1.21	$1.24^{\star}$	1.11
	12m	$1.38^{\star}$	$1.42^{\star\star}$	$1.39^{\star\star}$	$1.31^{\star\star}$	$1.29^{\star\star}$	$1.39^{\star}$	1.27	1.12	1.06	1.10
	24m	$1.19^{\star}$	$1.20^{\star}$	$1.21^{\star}$	$1.19^{\star}$	$1.26^{\star}$	1.11	1.07	1.03	1.08	1.05
	60m	0.99	1.01	1.01	0.99	0.92	0.77°	0.87	0.90	0.80	0.64°
Japan											
τ	1m	0.88	1.00	1.05	0.95	1.03	$0.52^{\circ\circ}$	$0.64^{\circ\circ}$	0.94	$0.88^{\circ\circ}$	0.90
	3m	0.94	1.03	1.04	1.04	1.08	$0.59^{\circ\circ}$	$0.68^{\circ\circ}$	0.91	0.97	1.00
	12m	1.04	1.03	1.02	1.07	1.12	0.89	0.92	0.99	1.00	1.02
	24m	1.00	1.00	1.00	0.99	1.01	1.00	1.00	1.00	1.00	1.00
	60m	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
United States											
τ	1m	0.93°	1.14	1.13	1.03	1.10	0.84°	$1.13^{\star}$	1.13	0.97	0.91
	3m	1.10	$1.22^{\star}$	$1.22^{\star}$	$1.22^{\star\star}$	$1.25^{\star\star}$	1.10	1.18	1.14	1.10	1.14
	12m	1.14	$1.18^{\star}$	$1.20^{\star}$	$1.26^{\star\star}$	$1.23^{\star\star}$	1.12	1.10	1.10	1.15	1.11
	24m	0.96	0.97	1.00	1.01	1.04	1.00	1.00	1.00	1.04	1.08
	60m	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99°	$0.81^{\circ\circ}$

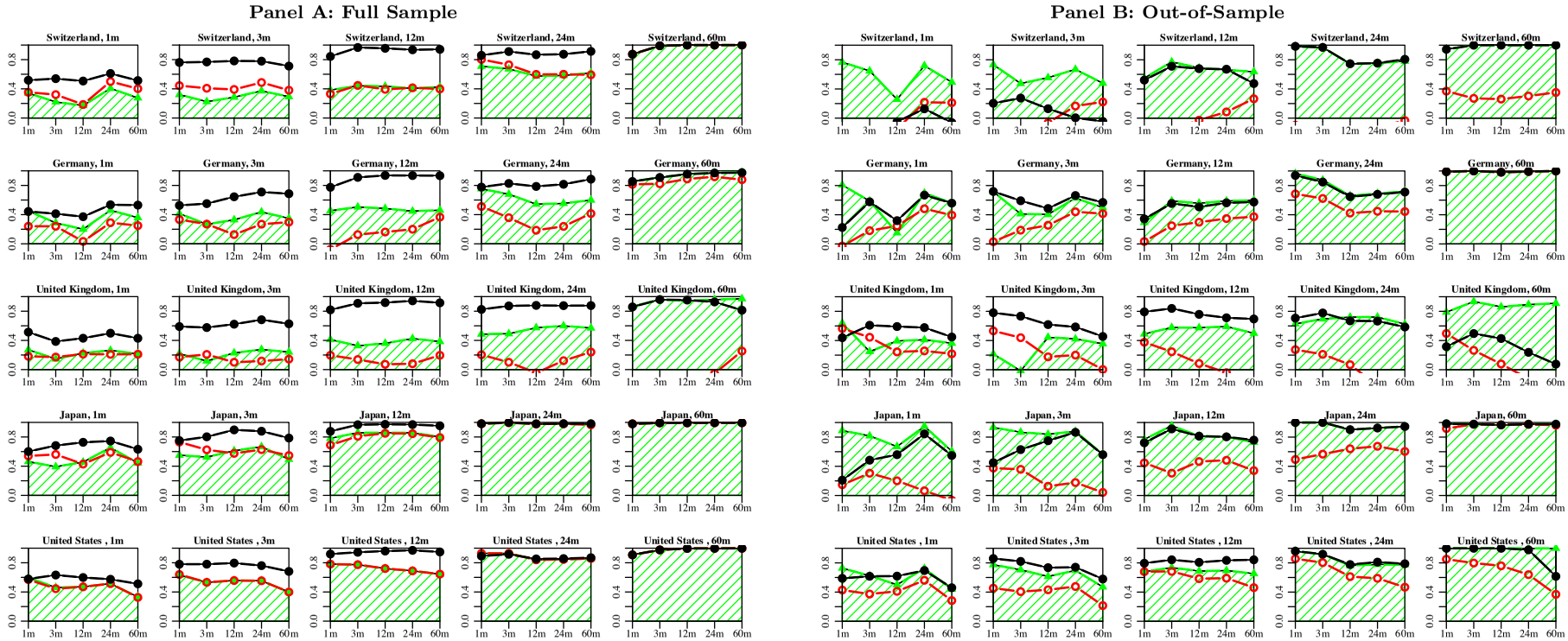
Notes: The Table presents measures of the directional accuracy of extended estimation compared to EH-consistent constant risk premium forecasts based on hit ratios that measuring the ratio of correctly signed forecasts. Values reported are computed as the fraction of the ATSM model hit ratio relative to the EH hit ratio. Estimates with a * indicate that the model hit ratio exceeds the 95%-percentile of the bootstrapped constant risk premium hit ratio distribution. ** indicates that the 5%-percentile of the model distribution is higher than the 95%-percentile of the constant risk premium hit ratio distribution. $^{\circ}$ and $^{\circ\circ}$ indicate analogous results when the model hit ratio is lower than the constant risk premium hit ratio. We present in- and out-of-sample results for 25 horizon (τ) and maturity ($T-\tau$) combinations, respectively. The horizons are indicated in the rows, the maturities in the columns. For the in-sample analysis, we estimate the models using monthly data from September 1989 for Japan, January 1988 for Switzerland, and April 1987 for Germany, UK, and US. The sample period ends in March 2011 for all countries. In the out-of-sample analysis, we estimate for each country data set the first model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t.

Figure IA.1: Average Excess Returns: International Evidence



Notes: The Figure plots averages of annualized excess returns of trading the longer-term bond with maturity T over a horizon τ. Each line represents the term structure of excess returns for a given horizon τ for maturities indicated on the x-axis. The left column represents the excess returns of a buy and hold strategy. The right columns plots excess returns of optimal bond portfolios of investors that have perfect foresight; for an investment horizon τ, an investor with target volatility of $\sigma^* = 2\%$ and relative risk aversion $\rho = 3$ optimally allocates his wealth between the risk-free bond with maturity τ and the risky bond with maturity $T > \tau$. The graphs are based on monthly data from September 1989 for Japan, January 1988 for Switzerland, and April 1987 for Germany, UK, and US. The sample period ends in March 2011 for all countries.

Figure IA.2: Excess Returns relative to Perfect Foresight: International Evidence



Notes: The Figure plots average excess returns of bond portfolios constructed based on EH and model forecasts relative to average excess returns of portfolios using perfect foresight forecasts. Portfolios using EH-consistent constant risk premium forecasts are represented by the green lines (with triangles and shaded area below). Portfolios based on model forecasts using the standard estimation procedure are plotted in red (with circles) and portfolios using extended estimation forecasts are displayed in black (with bullets). The underlying portfolios are generated by an investor with target volatility of $\sigma^* = 2\%$ and relative risk aversion $\rho = 3$ who optimally allocates his wealth between the risk-free bond with maturity τ and the risky bond with maturity $T > \tau$. For each country, we present separate graphs for five horizons τ, where the corresponding maturities are given on the x-axis that displays $T - \tau$. For the in-sample analysis (Panel A), we estimate the models using monthly data from September 1989 for Japan, January 1988 for Switzerland, and April 1987 for Germany, UK, and US. The sample period ends in March 2011 for all countries.In the out-of-sample analysis (Panel B), we estimate for each country data set the first model based on the first ten years of data available. Subsequently, we generate return forecasts every month t by re-estimating the model using only information available up to time t.