Keep it simple:
Dynamic bond portfolios under parameter uncertainty

Peter Feldhütter[a] Linda S. Larsen[b] Claus Munk[c] Anders B. Trolle[d]

April 19, 2012

Abstract: We empirically investigate the importance of parameter uncertainty to bond investors. Using a Bayesian approach, we quantify the expected utility loss due to parameter uncertainty from following seemingly optimal dynamic portfolio strategies. Expected utility losses are increasing in the number of term structure factors and the complexity of the risk premium specification. Even with long data sets to estimate parameters, an investor with typical risk aversion is better off following a portfolio strategy implied by a misspecified but parsimonious model than a correctly-specified but difficult-to-estimate three-factor affine model with time-varying risk premia.

Keywords: Suboptimal investments, parameter uncertainty, utility losses, bond portfolios, MCMC estimation

JEL subject codes: G11, G12, C13

[a] London Business School, Regent's Park, London NW1 4SA, United Kingdom. E-mail: pfeldhutter@london.edu

[b] Department of Business and Economics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. E-mail: lsn@sam.sdu.dk

[c] Department of Economics and Business & Department of Mathematics, Aarhus University, Bartholins Allé 10, DK-8000 Aarhus C, Denmark. E-mail: cmunk@econ.au.dk

[d] Swiss Finance Institute and Ecole Polytechnique Fédérale de Lausanne, Quartier UNIL-Dorigny, Extranef 216, CH-1015 Lausanne, Switzerland. E-mail: anders.trolle@epfl.ch

We are grateful for comments and suggestions from Christian Riis Flor, Simon Lysbjerg Hansen, Frank de Jong, and Carsten Sørensen.

Keep it simple:
Dynamic bond portfolios under parameter uncertainty

Abstract: We empirically investigate the importance of parameter uncertainty to bond investors. Using a Bayesian approach, we quantify the expected utility loss due to parameter uncertainty from following seemingly optimal dynamic portfolio strategies. Expected utility losses are increasing in the number of term structure factors and the complexity of the risk premium specification. Even with long data sets to estimate parameters, an investor with typical risk aversion is better off following a portfolio strategy implied by a misspecified but parsimonious model than a correctly-specified but difficult-to-estimate three-factor affine model with time-varying risk premia.

Keywords: Suboptimal investments, parameter uncertainty, utility losses, bond portfolios, MCMC estimation

JEL subject codes: G11, G12, C13

1 Introduction

The portfolio choice literature has explored a number of factors potentially relevant to investors: return predictability, stochastic volatility, stochastic correlation, jumps in stock prices, the housing market, and human capital, among others.[1] The broad conclusion is that these features are important for investors. These studies assume that investors know the correct model and model parameters. In this paper, we investigate the importance of parameter uncertainty and potential model misspecification to investors.

We focus on bond markets and estimate four term structure models by Markov Chain Monte Carlo (MCMC) on a panel data set of U.S. Treasury yields with daily observations over the period 1971-2006. The models are one- and three-factor affine models with both constant (completely affine) and time-varying (essentially affine) risk premia. We find that market price of risk parameters are imprecisely estimated; confidence intervals are wide and often parameters are not statistically significant. We determine the optimal dynamic investment strategy in all four models. Since the portfolio allocations are sensitive to market price of risk parameters, the allocations have wide confidence intervals as well. Often it is not even clear whether the investor should take a short or a long position in a given bond.

We investigate how the expected utility of the investor is affected by the large uncertainty in the market prices of risk by taking a Bayesian approach. In the first part of our empirical analysis, we assume the investor knows the true model but there is uncertainty regarding the estimated parameters. Assume that we know the true parameters of the model and that the investor's estimated parameters are different from the true ones. In this case, the investor's investment strategy is suboptimal, and we define the conditional utility loss as the fraction of initial wealth the investor would be willing to sacrifice to be able to follow the optimal strategy based on the true parameters. In practice, we do not know the true parameters, but our MCMC estimation gives us a posterior distribution of the true parameters. For a given portfolio strategy, we define the expected utility loss due to parameter uncertainty as the conditional utility loss integrated over the posterior distri-

[1]See e.g. Liu, Longstaff, and Pan (2003), Liu and Pan (2003), Cocco, Gomes, and Maenhout (2005), Yao and Zhang (2005), Chacko and Viceira (2005), Sangvinatsos and Wachter (2005), Liu (2007), Benzoni, Collin-Dufresne, and Goldstein (2007), Munk and Sørensen (2010), Buraschi, Porchia, and Trojani (2010), Koijen, Nijman, and Werker (2010), Kraft and Munk (2011), and Lynch and Tan (2011).

bution of the true parameters. The expected utility loss can be interpreted as the fraction of initial wealth the investor is willing to give up to live in a world with no parameter uncertainty.

We find that parameter uncertainty leads to large expected utility losses. What is crucial in determining the losses is whether risk premia are time-varying or not. If risk premia are time-varying, there is a significant probability of a future scenario of high expected excess returns in which modest deviations in the risk premium parameters lead to very different portfolio weights. In these scenarios the investor's portfolio weights are often off because of parameter uncertainty and this leads to high expected utility losses. For example, in the three-factor essentially affine model - the model used by Sangvinatsos and Wachter (2005) - an investor with a relative risk aversion of five and an investment horizon of five years has an expected utility loss of 66%. In contrast, the three-factor completely affine model only has an expected utility loss of 4.7%.

Although there are high expected utility losses due to parameter uncertainty in the three-factor essentially affine model, the model better captures time-variation in expected excess returns compared to the three simpler models (Dai and Singleton 2002, and Duffee 2002).[2] In the second part of the empirical analysis, we account for parameter uncertainty as in the first part, but we assume that the three-factor essentially affine model describes the true yield curve dynamics. Under this assumption, we calculate expected utility losses for investors basing their strategies on one of the three simpler models. By doing so, we quantify the tradeoff between the severity of parameter uncertainty and capturing the predictability in bond returns. In the three-factor essentially affine model the expected utility losses are due to parameter uncertainty, while in the three simpler models the losses are due to parameter uncertainty and model misspecification.

We find that long-term investors with moderate to high risk aversion are better off basing their portfolio decisions on more parsimonious models, since expected utility losses for these models due to both parameter uncertainty and model misspecification are smaller than the expected utility loss for the true model solely due to parameter uncertainty. As

[2] A number of recent papers argue that a component of the variation in bond risk premia is unspanned by the term structure; see, e.g. Joslin, Priebsch, and Singleton (2010) and Duffee (2011). Since the unspanned component is also associated with significant uncertainty, our results likely carry over to such an extended setting.

mentioned earlier, an investor with a relative risk aversion of five and an investment horizon of five years has an expected utility loss of 66% when basing his strategy on the "true" three-factor model with time-varying risk premia. Basing his strategy on a misspecified one-factor model with constant risk premia leads to an expected utility loss of only 54%, although there is model misspecification on top of parameter uncertainty. This is because the bond holdings are less extreme in the simple model and therefore the probability of having a poor performance is much smaller than in the "true" complex model. Hence, the suboptimal investment strategy based on the one-factor model with constant risk premia carries a 18% smaller utility loss compared to the investment strategy based on the "true" model. For a higher level of relative risk aversion or a longer investment horizon the difference in the average utility loss becomes even bigger.

For most of the paper, we compute expected utility losses assuming that the economy is in steady state. To see if there is significant variation over time, we also calculate expected utility losses where we condition on the current yield curve. For most of the sample period, the expected utility losses are higher for the three-factor essentially affine model relative to the simpler models. Interestingly, in the beginning of the 90s and the early part of the 00s the investor prefers to follow the three-factor model. In these periods the yield curve was steep and, as shown by Campbell and Shiller (1991), these are times when excess returns are high. Only an investor basing his portfolio choice on the three-factor essentially affine model can exploit the high bond risk premia during these episodes. Thus, the typical investor should not avoid the three-factor essentially affine model altogether, but rather know when to use it.

To the best of our knowledge, our paper is the first to study, within a realistic setting, the quantitative effects on expected utility of parameter uncertainty and the tradeoff between parameter uncertainty and model misspecification. Our results strongly suggest that investors should incorporate parameter uncertainty into their decision process. Klein and Bawa (1976) is one of the first papers to study the effect of parameter uncertainty on optimal portfolio strategies. Barberis (2000) incorporates parameter uncertainty in a setting with predictability in asset returns. Other papers include Barry and Brown (1985), Brennan (1998), Xia (2001), Maenhout (2004, 2006), Brandt, Goyal, Santa-Clara, and Stroud (2005), and Johannes, Korteweg, and Polson (2011). So far, the models have been very stylized because incorporating parameter uncertainty in realistic, dynamic settings quickly becomes difficult. From an investor's perspective, our results suggest that incorporating parameter uncertainty in existing models is more important than building new models that capture additional features of the investment universe. Consistent with this conclusion Sarno, Schneider, and Wagner (2012) find that three-factor affine models forecast poorly out-of-sample.

The remainder of the paper is organized in the following way. Section 2 sets up the modeling framework, and specifies the investment strategies. Section 3 discusses the data and the estimation procedure. Section 4 presents the results. Section 5 summarizes and concludes.

2 The general setup

In this section we specify our assumptions about the dynamics of the term structure of interest rates and the preferences of the investor. We derive the optimal investment strategy in affine models and solve for the conditional utility loss associated with suboptimal strategies.

2.1 The dynamic term structure model

We consider an arbitrage-free economy where trading takes place continuously in time. We assume that the term structure of interest rates follows an affine dynamic term structure model. More specifically, let r_t denote the instantaneous interest rate and assume that

$$r_t = \delta_0 + \delta_X' X_t, \tag{1}$$

where δ_0 is a constant, δ_X is an $m \times 1$ vector, and $X_t = (X_{1t}, X_{2t}, \ldots, X_{mt})'$ is an $m \times 1$ vector of state variables that follows the process

$$dX_t = \kappa\,(\theta - X_t)\,dt + \sigma_X\,dz_t \tag{2}$$

under the physical measure $\mathbb{P}$, where $z = (z_t)$ is a standard m-dimensional standard Brownian motion. The $m \times m$ constant matrix σ_X is assumed invertible and determines the variance-covariance matrix of the state variables over the next instant, $\sigma_X \sigma_X'$, κ is an invertible $m \times m$ matrix, and θ is an $m \times 1$ vector. Furthermore, the market price of risk associated with the shock process z is assumed to be affine in X,

$$\lambda_t = \lambda_0 + \lambda_X X_t, \tag{3}$$

where λ_0 is an $m \times 1$ vector and λ_X is an invertible $m \times m$ matrix. Following Duffee (2002), the model is said to be *completely affine* if $\lambda_X = 0$ and *essentially affine* otherwise. It follows from these assumptions that the dynamics of the state variables under the risk-neutral probability measure $\mathbb{Q}$ is

$$dX_t = \tilde{\kappa} \left(\tilde{\theta} - X_t \right) dt + \sigma_X \, dz_t^{\mathbb{Q}}, \tag{4}$$

where $z = (z_t^{\mathbb{Q}})$ is a standard Brownian motion under $\mathbb{Q}$ with $dz_t^{\mathbb{Q}} = dz_t + \lambda_t dt$. Furthermore, $\tilde{\kappa} = \kappa + \sigma_X \lambda_X$ and $\tilde{\theta} = \tilde{\kappa}^{-1} (\kappa\theta - \sigma_X \lambda_0)$.

As shown by Duffie and Kan (1996), the price of a zero-coupon bond maturing at T takes the form $P_t^T = P^T(t, X_t)$ where

$$P^T(t, X) = \exp\left\{ -A(T-t) - B(T-t)' X \right\},$$

and $A : [0, T] \to \mathbb{R}$, $B : [0, T] \to \mathbb{R}^m$ are solutions to the system of ordinary differential equations (ODEs):

$$\frac{\partial B(\tau)}{\partial \tau} = \delta_X - \tilde{\kappa}' B(\tau) \tag{5}$$

$$\frac{\partial A(\tau)}{\partial \tau} = B(\tau)' \tilde{\kappa} \, \tilde{\theta} - \frac{1}{2} B(\tau)' \, \sigma_X \, \sigma_X' \, B(\tau)' + \delta_0 \tag{6}$$

with the boundary conditions $A(0) = B(0) = 0$. For some model specifications these equations have explicit solutions; if not, the equations are easily solved numerically. The dynamics of the zero-coupon bond price with maturity T follows from Ito's lemma:

$$\frac{dP_t^T}{P_t^T} = \left(r_t - B(T-t)' \sigma_X \lambda_t \right) dt - B(T-t)' \sigma_X dz_t. \tag{7}$$

2.2 The investor

The investor can invest in an instantaneously risk-free asset, interpreted as short-term cash deposits, which yields the continuously compounded rate of return r_t. In addition, the investor can invest in $n < \infty$ zero-coupon bonds. We represent the investment strategy of the investor by the n-dimensional continuous-time process $\pi = (\pi_t)$, where $\pi_t = (\pi_{1t}, \pi_{2t}, \ldots, \pi_{nt})'$ is the vector of fractions of wealth ("portfolio weights") invested in n different zero-coupon bonds at time t. The remaining fraction of wealth $1 - \pi_t' \mathbf{1}$ is invested in the instantaneously risk-free asset. We ignore intermediate consumption and income other than financial returns and assume that the time-to-maturity of the bonds that the investor trades in are the same at all dates. Let τ_i, $i = 1, \ldots, n$, denote the time-to-maturity of the i'th bond in the portfolio. Further, let $\mathcal{B}$ be the $m \times n$ matrix with the i'th column representing the B-vector associated with the i'th zero-coupon bond, i.e.,

$$\mathcal{B} = (B(\tau_1), B(\tau_2), \ldots, B(\tau_n)) . \tag{8}$$

For notational simplicity we suppress the dependence of $\mathcal{B}$ on the maturities of the bonds. The volatility matrix of the n bonds is then $-\mathcal{B}'\sigma_X$. Given a positive initial wealth W_0 and an investment strategy π_t in these n zero-coupon bonds, the investor's wealth will satisfy the self-financing condition

$$dW_t = W_t \left[r_t - \pi_t' \mathcal{B}' \sigma_X \lambda_t \right] dt - W_t \pi_t' \mathcal{B}' \sigma_X \, dz_t. \tag{9}$$

For any other set of n non-redundant bonds (or other interest rate derivatives) there exists a portfolio with the same sensitivity towards the exogenous shocks as the portfolio π of the designated fixed-maturity bonds, and this equivalent portfolio is an easy transformation of π.

We assume that the investor maximizes expected utility of wealth at some future date T and that the utility function is of the CRRA type. The indirect utility is given as

$$J(W, X, t) = \sup_{(\pi_s)_{s \in [t,T]}} \begin{cases} \mathbb{E}_{W,X,t} \left[\frac{1}{1-\gamma} W_T^{1-\gamma} \right], & \gamma > 1, \\ \mathbb{E}_{W,X,t} \left[\ln W_T \right], & \gamma = 1, \end{cases} \tag{10}$$

where $\mathbb{E}_{W,X,t}$ denotes the expectation operator given $W_t = W$ and $X_t = X$ under the physical measure $\mathbb{P}$, and γ is the constant relative risk aversion parameter. The optimal investment strategy π^* is the one satisfying (10).

2.3 The optimal investment strategy

If the m-factor version of the model (1)-(3) is assumed correct and the investor knows the true parameters of this model, the technique applied by Liu (2007) and Sangvinatsos and Wachter (2005) leads to a semi-analytical expression for the optimal portfolio strategy and the associated expected utility. This involves trading in $n = m$ bonds. For completeness, we state the result here (see Appendix A for a proof):

Proposition 1 *When the investor assumes the m-factor model* (1)-(3) *is correct, the optimal investment strategy is to invest in $n = m$ different zero-coupon bonds according to the portfolio weights*

$$\begin{aligned}\pi^*(X,t) = \frac{1}{\gamma}\left(-\sigma_X'\mathcal{B}\right)^{-1}(\lambda_0 + \lambda_X X) \\ + \frac{\gamma-1}{\gamma}\left(\sigma_X'\mathcal{B}\right)^{-1}\sigma_X'\left(F_2(T-t) + \frac{1}{2}\left(F_3(T-t) + F_3(T-t)'\right)X\right)\end{aligned} \tag{11}$$

with the remaining wealth invested in short-term deposits. If the model assumed by the investor is the data-generating process, the expected utility generated by this strategy is

$$J(W,X,t) = \begin{cases} \frac{1}{1-\gamma}\left(W\, e^{F_1(T-t)+F_2(T-t)'X+\frac{1}{2}X'F_3(T-t)X}\right)^{1-\gamma}, & \gamma > 1, \\ \ln W + F_1(T-t) + F_2(T-t)'X + \frac{1}{2}X'F_3(T-t)X, & \gamma = 1. \end{cases} \tag{12}$$

Here F_1, F_2, and F_3 solve a system of ODEs stated in (21)–(23) *in Appendix A.*

The optimal investment strategy is composed of: a speculative portfolio (the first term in (11)) and a hedge portfolio (the second term in (11)). The hedge portfolio describes how the investor should optimally hedge against the changes in the investment opportunity set as a result of stochastic variation in the short rate and in the market prices of risk. The hedge portfolio consists of two components: the component involving F_2 is due to the stochastic variation in the short rate, whereas the component involving F_3 is due to the stochastic variation in the market price of risk vector. If market prices of risk are constant – that is, $\lambda_X = 0$ – then $F_3 = 0$ and the second component of the hedge portfolio disappears. See Sangvinatsos and Wachter (2005) for a more detailed discussion of the investment strategy.

2.4 Investment strategies

We will consider four different investment strategies:

(i) The investment strategy implied by a three-factor essentially affine model. This strategy is given by (11) with $m = n = 3$.

(ii) The investment strategy implied by a completely affine model. This strategy is given by (11) with $m = n = 3$ and $\lambda_X = 0$.

(iii) The investment strategy implied by a one-factor essentially affine model. This strategy is given by (11) with $m = n = 1$.

(iv) The investment strategy implied by a one-factor completely affine model. This strategy is given by (11) with $m = n = 1$ and $\lambda_X = 0$.

In the empirical section we will quantify the impact of parameter uncertainty. To do this, we need the expected utility of following an investment strategy implied by model i with parameter vector $\hat{\Theta}_i$ given that the data-generating model is model j with parameter vector Θ_j. Often, we will call $\hat{\Theta}_i$ for the investor's estimated parameters. The models i and j might be the same or it might be the case that model i is simpler than model j.

If the investment strategy is based on parameters different from the true parameters, the strategy is suboptimal. This is also the case if the investment strategy is based on a simpler model than the true model. Here "simpler" means that some parameters are set to zero or the number of factors is lower. In Appendix A we show that the expected utility generated by such a suboptimal investment strategy is given by

$$\hat{J}(W, X, t) = \begin{cases} \frac{1}{1-\gamma}\left(W\, e^{\hat{C}_1(T-t)+\hat{C}_2(T-t)'X+\frac{1}{2}X'\hat{C}_3(T-t)X}\right)^{1-\gamma}, & \gamma > 1, \\ \ln W + \hat{C}_1(T-t) + \hat{C}_2(T-t)'X + \frac{1}{2}X'\hat{C}_3(T-t)X, & \gamma = 1, \end{cases} \tag{13}$$

where the deterministic functions $\hat{C}_1$, $\hat{C}_2$, and $\hat{C}_3$ solve the system of ODEs stated in (28)–(30) in Appendix A. The functions $\hat{C}_1$, $\hat{C}_2$, and $\hat{C}_3$ depend both on the true parameters of the correct model and on the estimated parameters of the (possibly incorrect) model. We use a hat ($\hat{}$) when parameters are those estimated by the investor while there is no hat on the true parameters. Likewise, quantities that depend on estimated parameters wear a hat.

It might seem that there should be a hat on the latent factor X in (13) as well, since the factor is also estimated. However, Collin-Dufresne, Goldstein, and Jones (2008) show that any latent vector can - given the parameter values - be rotated into quantities that are observable from the data. For a one-factor model the latent X can be rotated into the short rate while for a three-factor model the state vector X can be rotated into the short rate, slope, and curvature of the yield curve. For convenience, we keep the latent notation, but one should think of them as the observable and model-independent quantities short rate, slope, and curvature and we will at times suppress the dependence on them.[3]

[3]In the empirical section we rotate the latent factor in the one-factor essentially affine model into the short rate before applying Proposition 2: $dr_t = \kappa\left(\hat{\theta} - r_t\right) dt + \sigma_X \delta_X\, dz_t$, where $\hat{\theta} = \delta_X \theta + \delta_0$ and the market price of risk is given by $\lambda = \left(\lambda_0 - \frac{\lambda_X}{\delta_X}\delta_0\right) + \frac{\lambda_X}{\delta_X}\, r$. For the three-factor essentially affine model

By definition, a suboptimal investment strategy will generate a lower level of expected utility than the optimal investment strategy. We define the conditional utility loss $L(\hat{\Theta}_i|X,\tau;\Theta_j)$ from following a suboptimal strategy implied by using model i with estimated parameters $\hat{\Theta}_i$ as the maximum fraction of initial wealth that the investor would sacrifice to be able to invest according to the optimal strategy implied by the data-generating model j with parameters Θ_j. The conditional utility loss depends on the current state X and the remaining investment horizon $\tau = T - t$ of the investor. By definition, the conditional utility loss solves

$$\hat{J}(W,X,t) = J\left(W\{1 - L(\hat{\Theta}_i|X,\tau;\Theta_j)\}, X, t\right). \tag{14}$$

Straightforward calculations using (12) and (13) prove the following proposition:

Proposition 2 *Assume that model i with parameter Θ_i is the data-generating process. The conditional utility loss generated by an investment strategy implied by model j with parameters $\hat{\Theta}_j$ is given by*

$$L(\hat{\Theta}_i|X,\tau;\Theta_j) = 1 - e^{\hat{C}_1(\tau)-F_1(\tau)+\left(\hat{C}_2(\tau)-F_2(\tau)\right)'X+\frac{1}{2}X'\left(\hat{C}_3(\tau)-F_3(\tau)\right)X} \tag{15}$$

for $\gamma \geq 1$ and F_1, F_2, F_3, $\hat{C}_1$, $\hat{C}_2$, and $\hat{C}_3$ are solutions to ODEs given in Appendix A.

In the following, we will often suppress the dependence of the loss on the state X and the investment horizon τ and simply write it as $L(\hat{\Theta}_j|\Theta_i)$. Obviously, $0 \leq L(\hat{\Theta}_j|\Theta_i) \leq 1$ and $L(\Theta_i|\Theta_i) = 0$.[4]

---

the empirical results are not very sensitive to - given the current yield curve - whether the state vector of the supposedly true model or the state vector of the supposedly false model is used. For simplicity, we have therefore chosen not to do the rotation in this case. For the completely affine models, the investment strategy is not dependent on X.

[4]In fact, for some strategies, some (sufficiently long) investment horizons, and some values of the parameter vector Θ_i, the loss can equal 100% if model i is an essentially affine model. This phenomenon has been discussed in some stochastic volatility models for derivatives pricing, where it is termed "moment explosion", cf. Andersen and Piterbarg (2007) and Keller-Ressel (2011). It is related to the concept of "Nirvana solutions" introduced by Kim and Omberg (1996) for a dynamic portfolio problem with a stochastic Sharpe ratio of the risky asset: for some parameter settings (including a relative risk aversion smaller than one) the investor can achieve an infinite expected utility. The mirror image is that some strategies are so bad that, for some parameter settings (including a relative risk aversion exceeding one), they will generate an expected utility of minus infinity. While interesting, it is beyond the scope of this paper to precisely characterize such "anti-Nirvana strategies".

3 Estimation

In this section, we describe how we estimate the four models we examine in the empirical section: the essentially and completely affine models with one or three factors.

3.1 Estimation procedure

To avoid overidentification we apply the parametrization of Dai and Singleton (2000) and assume that

(a) $\tilde{\theta} = \tilde{\kappa}^{-1} (\kappa\theta - \sigma_X \lambda_0) = 0$,

(b) σ_X equals the $(m \times m)$-identity matrix, and

(c) $\tilde{\kappa} = \kappa + \sigma_X \lambda_X$ is a $(m \times m)$-lower triangular matrix.

We adopt a Bayesian approach and estimate the models by Markov Chain Monte Carlo (MCMC) as proposed by Eraker (2001).[5] MCMC has also been used for estimating term structure models by, e.g., Ang, Dong, and Piazzesi (2007), Feldhütter (2008), Kaminska, Vayanos, and Zinna (2011), and Sarno, Schneider, and Wagner (2011). At time $t = 1, ..., T$ we observe k yields which are stacked in a k-vector

$$Y_t = (Y(t, \tau_1), ..., Y(t, \tau_k))'.$$

The yields are all observed with a measurement error

$$Y_t = A + BX_t + \epsilon_t,$$

where A is a k-vector and B is a $k \times m$ matrix. We assume that the measurement errors are independent and normally distributed with zero mean and common variance such that

$$\epsilon_t \sim N(0, D), \qquad D = \varphi^2 I_k,$$

where I_k denotes the $k \times k$ identity matrix. To simplify the notation in the following, we denote

$$\Theta^Q = (\tilde{\kappa}, \delta_0, \delta_X), \qquad \Theta^P = (\lambda_0, \lambda_X),$$

[5] See Robert and Casella (2004) for a general introduction to MCMC and Johannes and Polson (2009) for a survey of MCMC methods in financial econometrics.

and $\Theta = (\Theta^Q, \Theta^P, D)$.

We are interested in samples from the target distribution $p(\Theta, X\,|Y)$. The Hammersley-Clifford Theorem (Hammersley and Clifford, 1970, and Besag, 1974) implies that samples are obtained from the target distribution by sampling from a number of conditional distributions. Effectively, MCMC solves the problem of simulating from a complicated target distribution by simulating from simpler conditional distributions. If one samples directly from a full conditional distribution, the resulting algorithm is the Gibbs sampler (Geman and Geman, 1984). If it is not possible to sample directly from the full conditional distribution, one can sample by using the Metropolis-Hastings algorithm (Metropolis et al., 1953). We use a hybrid MCMC algorithm that combines the Gibbs sampler and the Metropolis-Hastings algorithm since not all the conditional distributions are known. Specifically, the MCMC algorithm is given by[6]

$$\begin{aligned}
p(\Theta^Q|\Theta^P, D, X, Y) &\sim \text{Metropolis-Hastings}\\
p(\lambda_0|\Theta_{\setminus\lambda_0}, X, Y) &\sim \text{Normal}\\
p(\lambda_X|\Theta_{\setminus\lambda_X}, X, Y) &\sim \text{Normal}\\
p(D|\Theta_{\setminus D}, X, Y) &\sim \text{Inverse Wishart}\\
p(X|\Theta, Y) &\sim \text{Metropolis-Hastings}
\end{aligned}$$

Details in the derivations of the conditionals and proposal distributions in the Metropolis-Hastings steps are given in Appendix B.1. Both the parameters and the latent processes are subject to constraints, and if a draw is violating a constraint it can simply be discarded (Gelfand et al., 1992).

In estimating each model we use an algorithm calibration period of 3 million draws, a burn-in period of 5 million draws, and an estimation period of 5 million draws. We keep every 5,000'th draw in the estimation period, which leaves 1,000 draws. For each of the four models we find our benchmark estimates among the 1,000 draws as follows. Following Collin-Dufresne et al. (2008), let ϕ_i denote the ith parameter draw and let $\tilde{\phi}_i$ denote the same vector normalized by the posterior standard deviations. The benchmark estimate is the draw i minimizing:

$$\sum_j \left|\tilde{\phi}_j - \tilde{\phi}_i\right|.$$

[6]Here $\Theta_{\setminus a}$ denotes the parameter vector excluding the parameter a.

This version of the multivariate posterior median ensures that parameter restrictions are satisfied for our parameter estimates, which might not be the case if the point estimates are based on univariate medians. For each parameter, we report the benchmark estimate along with a univariate confidence band based on the 2.5% and the 97.5% percentile of the 1,000 MCMC draws of the posterior distribution. Confidence bands for any quantity derived from the parameters are calculated as follows: for each parameter draw the quantity is calculated and a univariate confidence band is based on the 1,000 calculations of the quantity.

3.2 Data

We use daily (continuously compounded) 1-, 2-, 3-, 5-, 7-, and 10-year zero-coupon yields extracted from prices of off-the-run US Treasury securities for the period from August 16, 1971 to August 21, 2006. Off-the-run securities are defined as securities that are not among the two most recently issued securities with maturities of two, three, four, five, seven, and ten years. The data set is discussed in detail in Gürkaynak et al. (2006) and is posted on the website http://www.federalreserve.gov/pubs/feds/2006. Figure 1 depicts the time-series of the 1-, 5-, and 10-year yields.

[Figure 1 about here.]

4 Results

4.1 Parameter estimates

Tables 1, 2, and 3 display parameter estimates along with their confidence intervals for the four models considered in the paper. The market price of risk parameters λ_0 and λ_X are generally imprecisely estimated. For example, the estimate of λ_0 in the one-factor essentially affine model is 3.94 and the confidence band is from -0.33 to 9.18, so the confidence band is several times wider than the parameter estimate. For comparison, the estimate of δ_0 in the same model is -0.387 and the confidence band is from -0.390 to -0.385, so in this case the confidence band is tight. We also see that the confidence bands are much wider in the essentially affine models compared to the completely affine models. For example, the confidence band for λ_0 is 15 times wider in the one-factor essentially affine model compared to the completely affine one-factor model. Also, many of the risk premium parameters are not statistically significant. In the one-factor models two out of three are insignificant, while in the three-factor models nine out of 15 are insignificant. Dai and Singleton (2002) and Duffee (2002) simirlarly find that risk premium parameters are difficult to estimate accurately.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

4.2 Investment strategies

We now explore the size and statistical precision of portfolio weights implied by affine term structure models. As our benchmark case we consider an investor with an investment horizon of five years and a relative risk aversion of $\gamma = 5$. We consider the four investment strategies described in Section 2.4 and assume that the investor implements an investment strategy based on the estimated parameter values of the relevant model. We assume that the investor at every point in time can trade in a 1-year, 5-year, and 10-year zero-coupon bond as well as the instantaneously risk-free asset. Investors following the strategy implied by a one-factor model trade in only one bond, which we take to be the 5-year zero-coupon bond.[7] For now, we assume that the economy is in steady state and set the state vector in (11) equal to the unconditional mean under the actual measure $\mathbb{P}$.[8]

[Table 4 about here.]

Table 4 displays the four investment strategies for an investor with a relative risk aversion of $\gamma = 5$ and an investment horizon of 0, 5, and 10 years, respectively. Panel A displays for the three-factor essentially affine model the portfolio weights along with

[7] The return an investor earns by following a one-factor strategy may depend on the time-to-maturity of the bond, see e.g. Brennan and Xia (2002). However, the results are insensitive to which bond is used. These results are available on request.

[8] As mentioned earlier, the state vector can be viewed as the level, slope, and curvature of the yield curve. The unconditional mean can therefore be viewed as the average slope, level, and curvature (as defined in Collin-Dufresne et al. 2008) through the sample.

95% confidence bands at different investment horizons. The portfolio is highly levered consistent with the portfolio weights reported by Sangvinatsos and Wachter (2005). The confidence bands are wide and they suggest that it is not clear whether the investor should take a long or a short position in a given bond. For example, at a five-year investment horizon the investor has an estimated long position in the one-year bond of approximately 13 times his initial wealth. However, the 95% confidence interval goes from shorting the bond in an amount of 35 times his wealth to going long in the bond in an amount of 65 times his wealth.

Compared to the 12 risk premium parameters in the essentially affine three-factor model, the completely affine counterpart has only three. This leads to a reduced uncertainty in the portfolio weights; the confidence bands in Panel B are narrower and it is clear that the investor should borrow in the riskfree asset and go long in the one-year bond. Still, the portfolio weights in the completely affine case are also large, the sizes of the portfolio weights are statistically very uncertain, and for the five- and ten-year bonds the portfolio weights are not statistically different from zero.

For the one-factor models in Panel C in Table 4 we no longer see the extreme portfolio weights observed in the three-factor models. Furthermore, the confidence intervals are much narrower because of the more precise estimation of the parameters as well as the lower number of parameters used to calculate the strategy. The investment strategy in the three-factor essentially affine model involves 25 parameters, whereas only four parameters are used in the one-factor completely affine model. Even in this case, the portfolio weights are not statistically different from zero.

4.3 Utility losses due to parameter uncertainty

The large uncertainty in portfolio weights documented in the previous section clearly has an impact on the utility of investors. To investigate this impact we take a Bayesian approach. In this section we assume that the investor is basing his investment strategy on the correct data-generating model, but there is uncertainty about the parameters.

Recall that $L(\hat{\Theta}|\Theta)$ in (15) denotes the utility loss from following the strategy implied by the parameter $\hat{\Theta}$ conditional on the true parameter vector being Θ.[9] In this section we

[9] The dimensionality of $\hat{\Theta}$, Θ, and X are as follows. When we are calculating losses for the one-factor

suppress the dependence of $\hat{\Theta}_i$ and Θ_j on the specific models i and j, since the models i and j are the same. There is uncertainty regarding Θ and the MCMC estimation gives a posterior distribution of Θ given the data denoted $p(\Theta|Y)$. For each investment strategy we compute an expected utility loss by integrating the conditional utility loss over the posterior distribution of Θ_{true}:

$$\bar{L}(\hat{\Theta}) = \int L(\hat{\Theta}|\Theta)p(\Theta|Y)\, d\Theta. \tag{16}$$

We can interpret the expected utility loss as the fraction of wealth the investor is willing to give up to live in a world without parameter uncertainty. Korteweg and Polson (2009) use a similar approach to estimate the impact of parameter uncertainty on corporate credit spreads.

Table 5 illustrates the expected utility losses due to parameter uncertainty in the four models for different combinations of the risk aversion and the investment horizon. The expected utility losses are large in the three-factor essentially affine model. For example, Panel B shows that an investor with a risk aversion of $\gamma = 5$ and an investment horizon of five years is willing to give up 66% of his wealth to avoid parameter uncertainty. This compares to just 1.4% in the one-factor completely affine model. Thus, using a more complex model carries a larger utility loss because of parameter uncertainty.

[Table 5 about here.]

For any given model and a given risk aversion the expected utility loss is obviously increasing in the investment horizon. The impact of risk aversion on the utility loss depends on the relative parameter sensitivity of the speculative portfolio and the hedge portfolio. Table 5 shows that the utility loss is increasing in risk aversion over the 5- and 10-year horizons in the three-factor essentially affine model, whereas the utility loss is decreasing in risk aversion for the other models. Interestingly, the table shows that the utility loss due to parameter uncertainty cannot be proxied by either the number of parameters or the statistical uncertainty of the portfolio weights. The three-factor completely affine

completely affine model, both $\hat{\Theta}$ and Θ are 4×1 vectors, and X in (15) is the short rate. When we are calculating losses for the three-factor essentially affine model, both $\hat{\Theta}$ and Θ are 25×1 vectors, and X in (15) is a 3×1 vector containing the short rate, slope, and curvature of the yield curve. From this the dimensionality in the one-factor essentially affine model and three-factor completely affine model should be clear.

model has 24 parameters compared to the five in the one-factor essentially affine model, and the portfolio weights are more extreme and have larger confidence bands (when the economy is in steady state) as seen in Table 4. Still, the expected utility loss for $\gamma = 5$ and an investment horizon of five years is only 4.7% in the three-factor completely affine model compared to 11.3% in the one-factor essentially affine model. The reason is that the one-factor essentially affine model has time-varying risk premia. The model has modest portfolio weights in steady state, but when Treasury yields spike there are large expected excess returns and the portfolio weights become large and uncertain with possible large utility losses. In contrast, the three-factor completely affine model has constant portfolio weights over time.[10]

To further examine the difference in utility losses in the four models, Figure 2 shows the density of the utility losses for an investor with relative risk aversion $\gamma = 5$ and investment horizon $T = 5$. The expected utility losses in the second column of Panel B in Tabel 5 are the means in the distributions in the figure. We see that the combination of many risk parameters and time-varying risk premia leads to a high probability of seeing large losses in the three-factor essentially affine model; the probability of seeing utility losses of 95% or higher is 47%. In the one-factor essentially affine model there is also a tail of large losses, and the probability of utility losses of 95% or more is 3%. In contrast, the probability of seeing losses of 30% or more in the completely affine models is basically zero. The figure supports the conclusion that the large expected utility losses in the essentially affine models relative to completely affine models appear because there is a significant risk that the investor will suffer large losses.

[Figure 2 about here.]

Table 6 shows expected utility losses when we assume that market prices of risk are known with certainty.[11] We see that expected utility losses are miniscule and economically insignificant, so risk premium parameters determine almost exclusively expected utility losses. Market prices of risk are determined by the time series of yields, while the remaining

[10]For example, the confidence bands for the conditional portfolio weights in the one-factor essentially affine model are wider than in the three-factor completely affine model in the beginning of the eighties when yields were high. Results are available on request.

[11]Specifically, we set λ_0 and for the essentially affine models λ_X equal to their estimated values in all MCMC draws.

parameters are determined by both the time series and the cross-section of yields. Table 6 suggests that expected utility losses arising from parameter uncertainty are low for those parameters that are identified from the cross-section of yields, while they are large for those parameters that are identified only from the time series of yields.

[Table 6 about here.]

4.4 Utility losses due to parameter uncertainty and model misspecification

So far, we have shown that the three-factor essentially affine model produces investment strategies that are highly exposed to parameter uncertainty. However, empirical evidence suggests that the model captures the time-variation in risk premia in bonds, while the more parsimonious models we look at do not (Dai and Singleton 2002, and Duffee 2002). We now examine the trade-off between capturing the time-variation in the excess returns of bonds and minimizing the expected utility loss due to parameter uncertainty. We assume that the three-factor essentially affine model is the data-generating model and calculate expected utility losses in each of our four models. For the three-factor essentially affine model, expected utility losses are solely due to parameter uncertainty and identical to the losses calculated in the previous section. For our three other models, expected utility losses are due to both parameter uncertainty and model misspecification.

Table 7 shows the expected utility losses when we assume that the three-factor essentially affine model is the data-generating model. The losses for the three-factor essentially affine model is the same as those in Table 5 since there is no model misspecification in this case. The three other models are misspecified and we see that expected utility losses increase strongly. For the one-factor essentially affine model the expected utility losses go up by a factor 3-6, while in the completely affine models they go up by a factor 10-40. This is consistent with the finding of Sangvinatsos and Wachter (2005) that there are large utility gains from using a three-factor essentially affine model relative to a simpler model. However, the table also shows that for an investor with a relative risk aversion of $\gamma = 5$ and investment horizon of $T = 5$ years, the expected utility loss is 66% in the three-factor essentially affine model, while it is only 54% in the one-factor completely affine model, although there is model misspecification on top of parameter uncertainty. Hence, the sub-optimal investment strategy based on the one-factor model with constant risk premia in fact carries a 18% smaller utility loss compared to the investment strategy based on the true model. Even though the three-factor model captures time-variation in bond excess returns, the parameter uncertainty is so large that an investor would prefer to base his decisions on the one-factor model. As the table shows, this result holds even more strongly for higher relative risk aversions.

[Table 7 about here.]

Figure 3 shows the distribution of the utility losses in the four models when there is both parameter uncertainty and model misspecification. We find that there is a significant probability of seeing small losses in the three-factor essentially affine model. For example, the probability of utility losses smaller than 20% is 15% in the three-factor essentially affine model while it is small in the other three models.[12] So there is a significant probability of doing well in the three-factor essentially affine model relative to the other models. However, there is also a large probability mass for utility losses close to 100%. The probability of a utility loss of 95% or higher is 47% in the three-factor essentially affine model while it is around 3% in the other three models. Overall, there is a significant risk of doing poorly because of parameter uncertainty in the three-factor essentially affine model and this risk outweighs the chance of doing better compared to a simple model.

[Figure 3 about here.]

The expected utility losses reported so far have been computed assuming that the economy is in steady state. To study the time variation in expected utility losses due to parameter uncertainty and model misspecification, Figure 4 displays the time-series of expected utility losses for the three-factor essentially affine model and the one-factor completely affine model assuming an investment horizon of five years and a relative risk aversion of $\gamma = 5$. As we discussed previously, the latent state vector X can be rotated into the observable short rate, slope, and curvature of the yield curve (in the one-factor models the rotation is into the short rate). At each day during the sample, we use the X implied by the current yield curve to calculate expected utility losses. The parsimonious one-factor model consistently outperforms the complex three-factor model except for two episodes in

[12] Based on the 1,000 MCMC draws, the probability is 0 in the one-factor models, 0.1% in the three-factor completely affine model, and 14.7% in the three-factor essentially affine model.

the early 1990s and 2000s. Figure 1 shows that during those two periods the term structure was steeply upward-sloping, and it is well known from Campbell and Shiller (1991) and others that expected excess bond returns are high in such states. Only an investor basing his portfolio choice on the three-factor essentially affine model can exploit the high bond risk premia during these episodes.[13] During the first part of the 1980s the yield curve was inversed, so the slope was unusually low. This is exploitable in the three-factor model and therefore we also see higher expected utility losses in the one-factor model in this case. However, yields were high during this period, magnifying expected excess returns and the importance of parameter uncertainty, and therefore the expected utility losses also increase in the three-factor model. In this case, expected excess returns are high but increased parameter uncertainty accompanies them and investors prefer to stick to the simple model.

[Figure 4 about here.]

5 Conclusion

The sizeable recent literature on optimal investment strategies in various dynamic settings makes the courageous assumption that the true model and its parameters are known with certainty. There is no consensus about how to incorporate model and parameter uncertainty into the decision problem, and the proposed methods for doing so seem very difficult to implement in realistic, dynamic settings and might also be difficult to communicate to real-life investors. Alternatively, investors can look for investment strategies that are less sensitive to model and parameter uncertainty. Our approach allows the investor to quantitatively compare different models and weigh the cost of parameter uncertainty against the benefit of capturing return predictability.

Our examination of the US Treasury bond market suggests that most investors are better off estimating a relatively simple model and implementing the investment strategy derived from this model instead of estimating a more realistic and complex model and fol-

[13]In principle, the one-factor essentially affine model also exhibits time-varying bond risk premia. However, in this model the single state variable is highly correlated with the level of interest rates, which is much less informative about bond risk premia than the slope of the term structure. In fact, the expected utility loss for the one-factor essentially affine model is highly correlated with that of the one-factor completely affine model and, for that reason, is not displayed in Figure 4.

lowing the corresponding investment strategy. This is because the complex model involves a high degree of parameter uncertainty.

Our analysis focuses on models for interest rates and bond prices, but our approach can be used in other markets. For example, papers that emphasize time-varying equity risk premia often report large utility gains from following portfolio strategies that take equity return predictability into account. However, the implied portfolio weights are often extreme and very volatile (see, e.g., Brennan et al. 1997, Campbell and Viceira 1999, and Campbell et al. 2003), while at the same time equity return predictability is also associated with a large amount of uncertainty (see, e.g., Goyal and Welch 2008). Therefore, it is plausible that our results extend to the equity market. The extent to which our results hold in other markets is an interesting topic for future research.

A Proofs

A.1 Proof of Proposition 1

To solve for the optimal investment strategy we use the Dynamic Programming Approach suggested by Merton (1969, 1971, 1973). The Hamilton-Jacobi-Bellman (HJB) equation associated with the dynamic optimization problem is given by

$$\begin{aligned} 0 = \sup_{\pi \in \mathbb{R}^d} \Big\{ & J_t + J_W \left[r - \pi' \mathcal{B}' \sigma_X \lambda \right] W + \frac{1}{2} J_{WW} W^2 \pi' \mathcal{B}' \sigma_X \sigma_X' \mathcal{B} \pi \\ & + J_X' \kappa \left(\theta - X \right) + \frac{1}{2} \operatorname{tr} \left(J_{XX} \, \sigma_X \, \sigma_X' \right) - W \pi' \mathcal{B}' \sigma_X \sigma_X' J_{WX} \Big\} \end{aligned} \tag{17}$$

with the terminal condition $J(W, X, T) = \frac{W^{1-\gamma}}{1-\gamma}$ if $\gamma \neq 1$ and $J(W, X, T) = \ln W$ if $\gamma = 1$. The subscripts on J denote the partial derivatives.

The first order condition w.r.t. π implies that a candidate for the optimal investment strategy is given by

$$\pi^*(W, X, t) = \frac{J_W}{J_{WW} W} \left(\sigma_X' \mathcal{B} \right)^{-1} \lambda + \frac{1}{J_{WW} W} \left(\sigma_X' \mathcal{B} \right)^{-1} \sigma_X' J_{WX}. \tag{18}$$

By substituting the candidate for the optimal investment strategy into the HJB equation we get that

$$\begin{aligned} 0 = & J_t + J_W \left[r - (\pi^*)' \mathcal{B}' \sigma_X \eta \right] W + \frac{1}{2} J_{WW} W^2 (\pi^*)' \mathcal{B}' \sigma_X \sigma_X' \mathcal{B} \pi^* \\ & + J_X' \kappa \, (\theta - X) + \frac{1}{2} \operatorname{tr} \left(J_{XX} \, \sigma_X \, \sigma_X' \right) - W (\pi^*)' \mathcal{B}' \sigma_X \sigma_X' J_{WX}. \end{aligned} \tag{19}$$

An educated guess of the solution is

$$J(W, X, t) = \frac{1}{1-\gamma} \left(W e^{F_1(T-t) + F_2(T-t)' X + \frac{1}{2} X' F_3(\tau) X} \right)^{1-\gamma}. \tag{20}$$

The terminal condition of the HJB equation implies that $F_1(0) = F_2(0) = F_3(0) = 0$. Substituting the candidate for the optimal investment strategy (18) and the relevant derivatives of our guess into the HJB equation, simplifying, and finally matching coefficients on $X'[\,\cdot\,]X$, X', and the constant terms lead to the following system of ODEs:

$$\begin{aligned} -\frac{dF_1(\tau)}{d\tau} = & F_2(\tau)' \left[\kappa\theta + \frac{1-\gamma}{\gamma} \sigma_X \lambda_0 \right] + \frac{1-\gamma}{2\gamma} F_2(\tau)' \sigma_X \sigma_X' F_2(\tau) \\ & + \frac{1}{4} \operatorname{tr} \left(\left(F_3(\tau) + F_3(\tau)' \right) \sigma_X \sigma_X' \right) + \frac{1}{2\gamma} \lambda_0' \lambda_0 + \delta_0 \end{aligned} \tag{21}$$

$$\begin{aligned} -\frac{dF_2(\tau)}{d\tau} = & \left[\frac{1-\gamma}{\gamma} \lambda_X' \sigma_X' - \kappa' + \frac{1-\gamma}{2\gamma} \left(F_3(\tau) + F_3(\tau)' \right) \sigma_X \sigma_X' \right] F_2(\tau) \\ & + \frac{1}{2} \left(F_3(\tau) + F_3(\tau)' \right) \left[\kappa\theta + \frac{1-\gamma}{\gamma} \sigma_X \lambda_0 \right] + \frac{1}{\gamma} \lambda_X' \lambda_0 + \delta_X \end{aligned} \tag{22}$$

$$\begin{aligned} -\frac{dF_3(\tau)}{d\tau} = & \frac{1}{\gamma} \lambda_X' \lambda_X + \frac{1-\gamma}{\gamma} \lambda_X' \sigma_X' \left(F_3(\tau) + F_3(\tau)' \right) - \left(F_3(\tau) + F_3(\tau)' \right) \kappa \\ & + \frac{1-\gamma}{4\gamma} \left(F_3(\tau) + F_3(\tau)' \right) \sigma_X \sigma_X' \left(F_3(\tau) + F_3(\tau)' \right). \end{aligned} \tag{23}$$

Hence, our guess (20) is the solution to the HJB-equation if F_1, F_2, and F_3 solve the above system of ODEs. Finally, substituting the relevant derivatives of J into (18) gives the optimal strategy (11).

A.2 Proof of equation (13)

Assume that model j stated in (1)-(3) with parameter vector Θ_j is the data-generating model. Then any combination of an initial wealth and an investment strategy π will give rise to a terminal wealth W_T^π, and the expected utility associated with that is thus given by

$$\hat{J}(W,X,t) = \begin{cases} \mathbb{E}_t^{\mathbb{P}}\left[\frac{1}{1-\gamma}\left(W_T^\pi\right)^{1-\gamma}\right], & \gamma > 1, \\ \mathbb{E}_t^{\mathbb{P}}\left[\ln\left(W_T^\pi\right)\right], & \gamma = 1. \end{cases} \tag{24}$$

From Theorem 2 in Larsen and Munk (2012) it follows that the expected utility generated by the investment strategy, π, is given by

$$\hat{J}(W,X,t) = \begin{cases} \frac{1}{1-\gamma}\left(W\,e^{C(X,T-t)}\right)^{1-\gamma}, & \gamma > 1, \\ \ln W + C(X,T-t), & \gamma = 1, \end{cases} \tag{25}$$

where the function $C(X,\tau)$ satisfies the PDE

$$\begin{aligned} &-\frac{\partial C}{\partial \tau} + \left(\kappa\left(\theta - X\right) - (\gamma-1)\,\sigma_X\sigma_P'\pi(X,t)\right)'\frac{\partial C}{\partial X} + \frac{1}{2}\,\mathrm{tr}\left(\frac{\partial^2 C}{\partial X^2}\sigma_X\sigma_X'\right) \\ &-\frac{\gamma-1}{2}\left(\frac{\partial C}{\partial X}\right)'\sigma_X\sigma_X'\frac{\partial C}{\partial X} + r(X) + \pi(X,t)'\sigma_P\left[\lambda(X) - \frac{\gamma}{2}\sigma_P'\pi(X,t)\right] = 0 \end{aligned} \tag{26}$$

with the terminal condition $C(X,0) = 0$. $\sigma_P = \mathcal{B}'\sigma_X$ denotes the $n \times m$ volatility-matrix of the traded zero-coupon bonds. The suboptimal investment strategies we will consider can all be written on the form

$$\begin{aligned} \pi(X,t) = &\frac{1}{\gamma}\left(-\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\left(\hat{\lambda}_0 + \hat{\lambda}_X X\right) \\ &+ \frac{\gamma-1}{\gamma}\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\sigma}_X'\left(\hat{F}_2(T-t) + \frac{1}{2}\left(\hat{F}_3(T-t) + \hat{F}_3(T-t)'\right)X\right). \end{aligned} \tag{27}$$

where the hats ($\hat{}$) indicate terms that depend on the assumed parameter values. To follow notation, the investment strategy in (27) is based on model i with parameter vector $\hat{\Theta}_i$. For the specific suboptimal investment strategy stated above, an educated guess of a solution to the PDE is given by[14]

$$\hat{C}(X,\tau) = \hat{C}_1(\tau) + \hat{C}_2(\tau)'X + \frac{1}{2}X'\hat{C}_3(\tau)X.$$

Substituting in the relevant derivatives, the relevant investment strategy in (27), simplifying, and finally matching coefficients on $X'[\cdot]X$, X', and the constant terms lead to the following system of

[14]For the completely affine models the matrix function $C_3(\cdot)$ as well as $\hat{F}_3(\cdot)$ is put equal to zero.

ODEs:

$$
\begin{aligned}
-\frac{d\hat{C}_1(\tau)}{d\tau} &= \hat{C}_2(\tau)' \left[\kappa\theta + \frac{\gamma-1}{\gamma}\sigma_X\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\hat{\lambda}}_0\right] - \frac{\gamma-1}{2}\hat{C}_2(\tau)'\sigma_X\sigma_X'\hat{C}_2(\tau) \\
&\quad + \frac{\gamma-1}{\gamma}\hat{\hat{F}}_2(\tau)'\hat{\sigma}_X\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\left[\lambda_0 + \sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\hat{\lambda}}_0 - (\gamma-1)\sigma_X'\hat{C}_2(\tau)\right] \\
&\quad - \frac{(\gamma-1)^2}{2\gamma}\hat{\hat{F}}_2(\tau)'\hat{\sigma}_X\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\sigma}_X'\hat{\hat{F}}_2(\tau) \\
&\quad + \frac{1}{4}\operatorname{tr}\left(\left(\hat{C}_3(\tau) + \hat{C}_3(\tau)'\right)\sigma_X\sigma_X'\right) - \frac{1}{\gamma}\hat{\hat{\lambda}}_0'\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\lambda_0 \\
&\quad - \frac{1}{2\gamma}\hat{\hat{\lambda}}_0'\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\hat{\lambda}}_0 + \delta_0
\end{aligned}
\tag{28}
$$

$$
\begin{aligned}
-\frac{d\hat{C}_2(\tau)}{d\tau} &= \left[\frac{\gamma-1}{\gamma}\hat{\hat{\lambda}}_X'\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\sigma_X' - \kappa' - \frac{\gamma-1}{2}\left(\hat{C}_3(\tau) + \hat{C}_3(\tau)'\right)\sigma_X\sigma_X'\right]\hat{C}_2(\tau) \\
&\quad + \frac{1}{2}\left(\hat{C}_3(\tau) + \hat{C}_3(\tau)'\right)\left[\kappa\theta + \frac{\gamma-1}{\gamma}\sigma_X\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\left(\hat{\hat{\lambda}}_0 - (\gamma-1)\,\hat{\sigma}_X'\hat{\hat{F}}_2(\tau)\right)\right] \\
&\quad + \frac{\gamma-1}{\gamma}\left[\lambda_X'\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\sigma}_X' + \hat{\hat{\lambda}}_X'\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\sigma}_X'\right]\hat{\hat{F}}_2(\tau) \\
&\quad + \frac{\gamma-1}{2\gamma}\left(\hat{\hat{F}}_3(\tau) + \hat{\hat{F}}_3(\tau)'\right)\hat{\sigma}_X\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\left[\lambda_0 + \sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\hat{\lambda}}_0\right] \\
&\quad - \frac{(\gamma-1)^2}{2\gamma}\left(\hat{\hat{F}}_3(\tau) + \hat{\hat{F}}_3(\tau)'\right)\hat{\sigma}_X\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\left[\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\sigma}_X'\hat{\hat{F}}_2(\tau) + \sigma_X'\hat{C}_2(\tau)\right] \\
&\quad - \frac{1}{\gamma}\lambda_X'\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\hat{\lambda}}_0 - \frac{1}{\gamma}\hat{\hat{\lambda}}_X'\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\lambda_0 \\
&\quad - \frac{1}{\gamma}\hat{\hat{\lambda}}_X'\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\hat{\lambda}}_0 + \delta_X
\end{aligned}
\tag{29}
$$

$$
\begin{aligned}
-\frac{d\hat{C}_3(\tau)}{d\tau} &= -\frac{1}{\gamma}\hat{\hat{\lambda}}_X'\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\left(2\lambda_X + \sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\hat{\lambda}}_X\right) \\
&\quad + \frac{\gamma-1}{\gamma}\lambda_X'\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\sigma}_X\left(\hat{\hat{F}}_3(\tau) + \hat{\hat{F}}_3(\tau)'\right) \\
&\quad + \frac{\gamma-1}{\gamma}\hat{\hat{\lambda}}_X'\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\sigma_X'\left(\hat{C}_3(\tau) + \hat{C}_3(\tau)'\right) \\
&\quad + \frac{\gamma-1}{\gamma}\hat{\hat{\lambda}}_X'\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\sigma}_X'\left(\hat{\hat{F}}_3(\tau) + \hat{\hat{F}}_3(\tau)'\right) \\
&\quad - \left(\hat{C}_3(\tau) + \hat{C}_3(\tau)'\right)\kappa - \frac{\gamma-1}{4}\left(\hat{C}_3(\tau) + \hat{C}_3(\tau)'\right)\sigma_X\sigma_X'\left(\hat{C}_3(\tau) + \hat{C}_3(\tau)'\right) \\
&\quad - \frac{(\gamma-1)^2}{4\gamma}\left(\hat{\hat{F}}_3(\tau) + \hat{\hat{F}}_3(\tau)'\right)\hat{\sigma}_X\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\sigma_P'\left(\hat{\sigma}_X'\hat{\mathcal{B}}\right)^{-1}\hat{\sigma}_X'\left(\hat{\hat{F}}_3(\tau) + \hat{\hat{F}}_3(\tau)'\right) \\
&\quad - \frac{(\gamma-1)^2}{2\gamma}\left(\hat{\hat{F}}_3(\tau) + \hat{\hat{F}}_3(\tau)'\right)\hat{\sigma}_X\left(\hat{\mathcal{B}}'\hat{\sigma}_X\right)^{-1}\sigma_P\sigma_X'\left(\hat{C}_3(\tau) + \hat{C}_3(\tau)'\right).
\end{aligned}
\tag{30}
$$

with boundary condition $\hat{C}_1(0) = \hat{C}_2(0) = \hat{C}_3(0) = 0$. The hats ($\hat{}$) on the parameters imply that the benchmark parameter estimates for the models should be used. In the case where there is only uncertainty regarding the parameter vector Θ and not the data-generating process, i.e. model i and j are the same, we have that $\hat{\hat{\lambda}}_X = \hat{\lambda}_X$, $\hat{\hat{\lambda}}_0 = \hat{\lambda}_0$, $\hat{\hat{F}}_2(\cdot) = \hat{F}_2(\cdot)$, and $\hat{\hat{F}}_3(\cdot) = \hat{F}_3(\cdot)$, that is, use the benchmark estimates for the three investment strategies. This is also the case if the data-generating process equals the three-factor essentially affine model, whereas the investor bases his investment strategy on one of the two completely affine models. However, if the investor bases his investment strategy on the one-factor essentially affine model, then

$$\hat{\hat{\lambda}}_X = \frac{\hat{\lambda}_X}{\hat{\delta}_X}\delta_X', \qquad \hat{\hat{\lambda}}_0 = \hat{\lambda}_0 - \frac{\hat{\lambda}_X}{\hat{\delta}_X}\left(\hat{\delta}_0 - \delta_0\right),$$
$$\hat{\hat{F}}_3(\tau) = \frac{\hat{F}_3(\tau)}{\hat{\delta}_X}\delta_X', \qquad \hat{\hat{F}}_2(\tau) = \hat{F}_2(\tau) - \frac{\hat{F}_3(\tau)}{\hat{\delta}_X}\left(\hat{\delta}_0 - \delta_0\right).$$

Note that the functions $\hat{F}_2(\tau)$ and $\hat{F}_3(\tau)$ solve the system of ODEs (22)–(23) from the optimal setup with the benchmark parameter estimates for each of the four models. Hence, our guess is the solution to the PDE (26) if $\hat{C}_1$, $\hat{C}_2$, and $\hat{C}_3$ solve the above system of ODEs.[15]

B Details of the MCMC estimation

First, the conditionals mentioned in the text are derived, and thereafter practical issues regarding the MCMC sampler are discussed.

B.1 Conditional Distributions

B.1.1 The Conditionals $p(X|\Theta)$ and $p(Y|\Theta, X)$

The conditional $p(X|\Theta)$ is used in several steps of the MCMC procedure and is calculated as

$$p(X|\Theta) = \left(\prod_{t=1}^{T} p(X_t|X_{t-1}, \Theta)\right) p(X_0).$$

The continuous-time specification in (4) is approximated using an Euler scheme[16]

$$X_{t+1} = X_t + \mu_{Xt}^{\mathbb{P}}\Delta_t + \sqrt{\Delta_t}\,\xi_{t+1},$$

where $\xi_{t+1} \sim N(0, I_N)$, Δt is the time between two observations, and $\mu_{Xt}^{\mathbb{P}}$ is the drift under $\mathbb{P}$. Therefore,

$$p(X|\Theta) \propto \exp\left\{-\frac{1}{2\Delta_t}\sum_{t=1}^{T}\sum_{i=1}^{3}\left[X_t - X_{t-1} - \mu_{Xt-1}^{\mathbb{P}}\Delta_t\right]_i^2\right\} p(X_0).$$

[15] Note that for the completely affine models we have that $F_3(\cdot) = \hat{C}_3(\cdot) = 0$ and hence the system of ODEs can be simplified significantly.

[16] The Euler scheme introduces some discretization error which may induce bias in the parameter estimates. This possible bias can be reduced using Tanner and Wong (1987)'s data augmentation scheme. However, the discretization bias is likely to be small for daily data.

If the difference between the actual yields and the model-implied yields at time t is denoted by $\hat{e}_t = Y_t - (A(\Theta) + B(\Theta)X_t)$, the density $p(y|\Theta, X)$ can be written as

$$p(Y|\Theta, X) \propto \prod_{i=1}^{k} \left(D_{ii}^{-\frac{T}{2}} \exp\left\{ -\frac{1}{2D_{ii}} \sum_{t=1}^{T} \hat{e}_{t,i}^2 \right\} \right) \propto \varphi^{-kT} \exp\left\{ -\frac{1}{2\varphi^2} \sum_{t=1}^{T} \hat{e}_t' \hat{e}_t \right\}.$$

B.1.2 The Hybrid MCMC algorithm

According to Bayes' theorem, the conditional of the risk premium parameters is given as

$$\begin{aligned} p(\lambda_0, \lambda_X | \Theta_{\backslash \lambda_0, \lambda_X}, X, Y) &\propto p(Y|\Theta, X)\, p(\lambda_0, \lambda_X | \Theta_{\backslash \lambda_0, \lambda_X}, X) \\ &\propto p(X|\Theta)\, p(\lambda_0, \lambda_X | \Theta_{\backslash \lambda_0, \lambda_X}), \end{aligned}$$

where $\Theta_{\backslash \lambda_0, \lambda_X}$ denotes the parameter vector without the parameters λ_0 and λ_X. We assume that the priors are a priori independent and in order to let the data dominate the results a standard diffuse, non-informative prior is adopted so $p(\lambda_0, \lambda_X | \Theta_{\backslash \lambda_0, \lambda_X}, X, Y) \propto p(X|\Theta)$ and λ_0, λ_X can be Gibbs sampled one column at a time from a multivariate normal distribution. The conditionals of the other model parameters are given as

$$p(\Theta_j | \Theta_{\backslash \Theta_j}, X, Y) \propto p(Y|\Theta, X)\, p(\Theta_j | \Theta_{\backslash \Theta_j}, X) \propto p(Y|\Theta, X)\, p(X|\Theta). \tag{31}$$

Equation (31) implies that the conditional of the variance of the measurement errors is given as

$$p(D|\Theta_{\backslash D}, X, Y) \propto p(Y|\Theta, X)\, p(D|\Theta_{\backslash D}).$$

The parameter φ^2 can therefore be Gibbs sampled from the inverse Wishart distribution, $\varphi^2 \sim IW(\sum_{t=1}^{T} \hat{e}_t' \hat{e}_t, kT)$.

To sample $\tilde{\kappa}, \delta_0$, and δ_X we use the Random Walk Metropolis-Hastings algorithm (RW-MH). Equation (31) gives the general expression for the conditional distribution.

The latent processes are sampled by sampling X_t, $t = 0, ..., T$ one at a time using the RW-MH procedure. For $t = 1, ..., T-1$ the conditional of X_t is given as

$$\begin{aligned} p(X_t | X_{\backslash t}, \Theta, Y) &\propto p(X_t | X_{t-1}, X_{t+1}, \Theta, Y_t) \\ &\propto p(Y_t | X_t, \Theta)\, p(X_t | X_{t-1}, X_{t+1}, \Theta) \\ &\propto p(Y_t | X_t, \Theta)\, p(X_t | X_{t-1}, \Theta)\, p(X_{t+1} | X_t, \Theta). \end{aligned}$$

For $t = 0$ the conditional is

$$p(X_0 | X_1, \Theta, Y) \propto p(X_1 | X_0, \Theta, Y)\, p(X_0) \propto p(X_1 | X_0, \Theta)\, p(X_0),$$

while for $t = T$ the conditional is

$$\begin{aligned} p(X_T|X_{\backslash X_T}, \Theta, Y) &\propto p(X_T|X_{T-1}, \Theta, Y) \\ &\propto p(Y_T|X_T, X_{T-1}, \Theta, Y_{\backslash Y_T})\, p(X_T|X_{T-1}, \Theta, Y_{\backslash Y_T}) \\ &\propto p(Y_T|X_T, \Theta)\, p(X_T|X_{T-1}, \Theta). \end{aligned}$$

The efficiency of the RW-MH algorithm depends crucially on the variance of the proposed normal distribution. If the variance is too low, the Markov chain will accept nearly every draw and converge very slowly, while it will reject a too high portion of the draws if the variance is too high. We therefore do an algorithm calibration and adjust the variance in the first eight million draws in the MCMC algorithm. Within each parameter block the variance of the individual parameters is the same, while across parameter blocks the variance may be different. Roberts et al. (1997) recommend acceptance rates close to $\frac{1}{4}$ for models of high dimension and therefore the standard deviation during the algorithm calibration is chosen as follows: Every 100'th draw the acceptance ratio of the parameters in a block is evaluated. If it is less than 5% the standard deviation is doubled, while if it is more than 40% it is cut in half. This step is prior to the burn-in period since the convergence results of RW-MH only apply if the variance is constant (otherwise the Markov property of the chain is lost). In estimating each model we use an algorithm calibration period of 3 million draws, where the variances of the normal proposal distributions are set, a burn-in period of 5 million draws, and an estimation period of 5 million draws. We keep every 5,000'th draw in the estimation period, which leaves 1,000 draws. For each parameter, we report point estimates along with univariate confidence bands based on the 2.5% and the 97.5% percentiles of the MCMC draws of the posterior distribution. As in Collin-Dufresne et al. (2008) we find point estimates as follows. Let ϕ_i denote the ith parameter draw and $\tilde{\phi}_i$ denote the same vector normalized by the posterior standard deviations. The point estimate is the draw i minimizing:

$$\sum_j \left| \tilde{\phi}_j - \tilde{\phi}_i \right|.$$

This version of the multivariate posterior median ensures that parameter restrictions are satisfied for our parameter estimates, which might not be the case if the point estimates are based on univariate medians.

All random numbers in the estimation are draws from Matlab 7.0's generator, which is based on Marsaglia and Zaman (1991)'s algorithm. The generator has a period of almost 2^{1430} and therefore the number of random draws in the estimation is not anywhere near the period of the random number generator.

References

Andersen, L. and V. Piterbarg (2007). Moment explosions in stochastic volatility models. *Finance and Stochastics 11*(1), 29–50.

Ang, A., S. Dong, and M. Piazzesi (2007). No-arbitrage taylor rules. Working paper.

Barberis, N. (2000). Investing for the Long Run when Returns are Predictable. *Journal of Finance 55*, 225–264.

Barry, C. B. and S. J. Brown (1985). Differential Information and Securtiy Market Equilibrium. *Journal of Financial and Quantitative Analysis 20*, 407–422.

Benzoni, L., P. Collin-Dufresne, and R. S. Goldstein (2007). Portfolio Choice over the Life-Cycle when the Stock and Labor Markets are Cointegrated. *Journal of Finance 62*(5), 2123–2167.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Association Series B 36*, 192–236.

Brandt, M. W., A. Goyal, P. Santa-Clara, and J. R. Stroud (2005). Simulation Approach to Dynamic Portfolio Choice with an Application to Learning About Return Predictability. *Review of Financial Studies 18*, 831–873.

Brennan, M. J. (1998). The Role of Learning in Dynamic Portfolio Decisions. *European Finance Review 1*(3), 295–306.

Brennan, M. J., E. S. Schwartz, and R. Lagnado (1997). Strategic Asset Allocation. *Journal of Economic Dynamics and Control 21*, 1377–1403.

Brennan, M. J. and Y. Xia (2002). Dynamic Asset Allocation under Inflation. *Journal of Finance 57*, 1201–1238.

Buraschi, A., P. Porchia, and F. Trojani (2010). Correlation Risk and Optimal Portfolio Choice . *Journal of Finance 65*(1), 393–420.

Campbell, J. Y., Y. L. Chan, and L. M. Viceira (2003). A Multivariate Model of Strategic Asset Allocation. *Journal of Financial Economics 67*, 41–80.

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

Campbell, J. Y. and L. M. Viceira (1999). Consumption and Portfolio Decisions when Expected Returns are Time-Varying. *The Quarterly Journal of Economics 114*, 433–495.

Chacko, G. and L. Viceira (2005). Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Markets. *Review of Financial Studies 18*(4), 1369–1402.

Cocco, J. F., F. J. Gomes, and P. J. Maenhout (2005). Portfolio Consumption and Portfolio Choice over the Life Cycle. *Review of Financial Studies 18*(2), 491–533.

Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones (2008). Identification of Maximal Affine Term Structure Models. *Journal of Finance 63*, 743–795.

Dai, Q. and K. Singleton (2000). Specification Analysis of Affine Term Structure Models. *Journal of Finance 55*, 1943–1978.

Dai, Q. and K. Singleton (2002). Expectation Puzzles, Time-varying Risk Premia, and Dynamic Models of the Term Structure. *Journal of Financial Economics 3*, 415–441.

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

Duffee, G. R. (2002). Term Premia and Interest Rate Forecast in Affine Models. *Journal of Finance 57*, 405–443.

Duffie, D. and R. Kan (1996). A Yield-factor Model of Interest Rates. *Mathematical Finance 6*, 379–406.

Eraker, B. (2001). MCMC Analysis of Diffusion Models with Application to Finance. *Journal of Business and Economic Statistics 19*, 177–191.

Feldhütter, P. (2008). Can Affine Models Match the Moments in Bond Yields? Working paper.

Gelfand, A. E., A. F. Smith, and T.-M. Lee (1992). Bayesian Analysis of Constrained Parameter and Truncated Data problems Using Gibbs Sampling. *Journal of the American Statistical Association 87*, 523–532.

Geman, S. and D. Geman (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Trans. on Pattern Analysis and Machine Intelligence 6*, 721–741.

Goyal, A. and I. Welch (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies 21*, 1455–1508.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2006). The U.S. Treasury Yield Curve: 1961 to the Present. Working paper, Divisions of Research and Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.

Hammersley, J. and P. Clifford (1970). Markov Fields on Finite Graphs and Lattices. Working paper.

Johannes, M., A. Korteweg, and N. Polson (2011). Sequential Learning, Predictive Regressions, and Optimal Portfolio Returns. Working paper.

Johannes, M. and N. Polson (2009). MCMC Methods for Financial Econometrics. In *Handbook of Financial Econometrics*, Volume 2, pp. 1–72. North-Holland.

Joslin, S., M. Priebsch, and K. J. Singleton (2010). Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks. Working paper.

Kaminska, I., D. Vayanos, and G. Zinna (2011). Preferred-Habitat Investors and the US Term Structure of Real Rates. Working paper.

Keller-Ressel, M. (2011). Moment explosions and long-term behavior of affine stochastic volatility models. *Mathematical Finance 21*(1), 73–98.

Kim, T. S. and E. Omberg (1996). Dynamic Nonmyopic Portfolio Behavior. *Review of Financial Studies 9*, 141–161.

Klein, R. W. and V. S. Bawa (1976). The Effect of Estimation Risk on Optimal Portfolio Choice. *Journal of Financial Economics 3*, 215–231.

Koijen, R. S. J., T. Nijman, and B. J. M. Werker (2010). When Can Life-Cycle Investors Benefit from Time-Varying Bond Risk Premia? *Review of Financial Studies 23*, 741–780.

Korteweg, A. G. and N. Polson (2009). Corporate Credit Spreads under Parameter Uncertainty. Working paper.

Kraft, H. and C. Munk (2011). Optimal Housing, Consumption, and Investment Decisions over the Life-Cycle. *Management Science 57*(6), 1025–1041.

Larsen, L. S. and C. Munk (2012). The Costs of Suboptimal Dynamic Asset Allocation:

General Results and Applications to Interest Rate Risk, Stock Volatility Risk and Growth/Value Tilts. *Journal of Economic Dynamics and Control 36*, 266–293.

Liu, J. (2007). Portfolio Selection in Stochastic Environments. *Review of Financial Studies 20*, 1–39.

Liu, J., F. Longstaff, and J. Pan (2003). Dynamic Asset Allocation with Event Risk. *Journal of Finance 58*(1), 231–259.

Liu, J. and J. Pan (2003). Dynamic Derivatives Strategies. *Journal of Financial Economics 69*, 401–430.

Lynch, A. and S. Tan (2011). Labor Income Dynamics at Business-cycle Frequencies: Implications for Portfolio Choice. *Journal of Financial Economics 101*, 333–359.

Maenhout, P. J. (2004). Robust Portfolio Rules and Asset Pricing. *Review of Financial Studies 17*, 951–983.

Maenhout, P. J. (2006). Robust Portfolio Rules and Detection-Error Probabilities for a Mean-Reverting Risk Premium. *Journal of Economic Theory 128*, 136–163.

Marsaglia, G. and A. Zaman (1991). A New Class of Random Number Generators. *Annals of Applied Probability 3*, 462–480.

Merton, R. C. (1969). Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case. *Rewiew of Economics and Statistics 51*, 247–257.

Merton, R. C. (1971). Optimum Consumption and Portfolio Rules in a Continuous-time Model. *Journal of Economic Theory 3*, 373–413.

Merton, R. C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica 41*, 867–887.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953). Equations of State Calculations by fast Computing Machines. *Journal of Chemical Physics 21*, 1087–1091.

Munk, C. and C. Sørensen (2010). Dynamic Asset Allocation with Stochastic Income and Interest Rates. *Journal of Financial Economics 96*, 433–462.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods* (Second ed.). Springer-Verlag.

Roberts, G., A. Gelman, and W. Gilks (1997). Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *Annals of Applied Probability 7*, 110–120.

Sangvinatsos, A. and J. A. Wachter (2005). Does the Failure of the Expectations Hypotheses Matter for Long-Term Investors? *Journal of Finance 60*, 179–230.

Sarno, L., P. Schneider, and C. Wagner (2011). Properties of Foreign Exchange Risk Premiums. *Journal of Financial Economics*, forthcoming.

Sarno, L., P. Schneider, and C. Wagner (2012). The Economic Value of Predicting Bond Risk Premia: Can Anything Beat the Expectations Hypothesis? Working paper.

Tanner, M. and W. Wong (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association 82*, 528–549.

Xia, Y. (2001). Learning about Predictatbility: The Effects of Parameter Uncertainty on Dynamic Asset Allocation. *Journal of Finance 56*, 205–244.

Yao, R. and H. H. Zhang (2005). Optimal Consumption and Portfolio Choices with Risky Housing and Borrowing Constraints. *Review of Financial Studies 18*, 197–239.

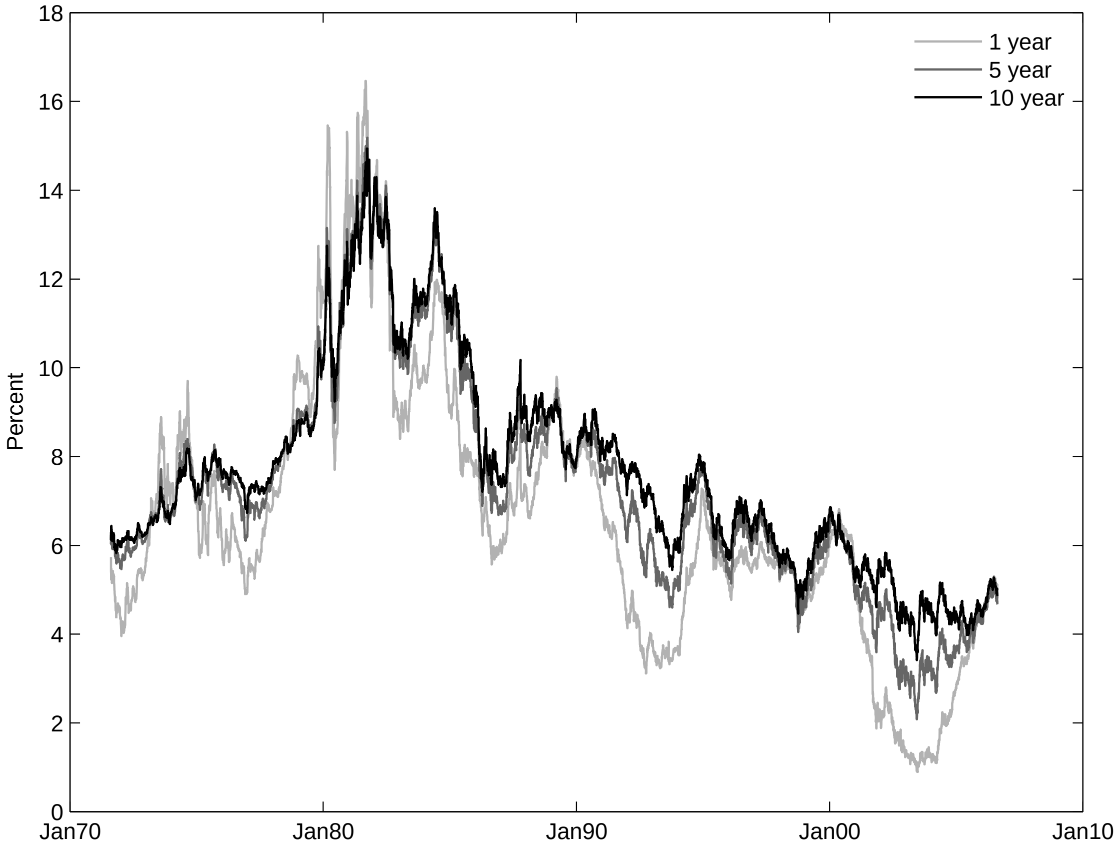

Figure 1: **Time-series of 1-, 5-, and 10-year zero-coupon Treasury yields.** The MCMC estimation is based on a panel data set of daily 1-, 2-, 3-, 5-, 7-, and 10-year zero-coupon Treasury yields from August 16, 1971 to August 21, 2006. The figure shows the time series of the 1-, 5-, and 10-year yield.

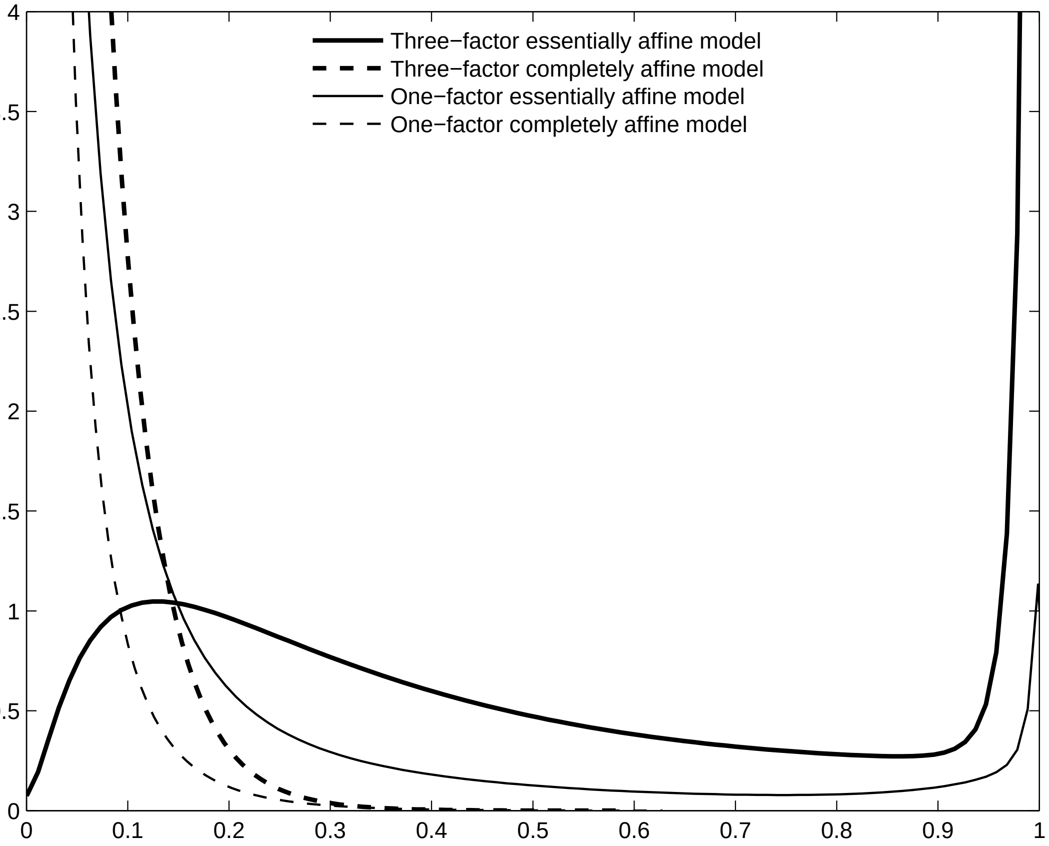

Figure 2: **Density of utility losses under parameter uncertainty.** For each of the four models in the figure it is assumed that the model is the data-generating model. The investor chooses a portfolio strategy implied by his model parameter estimates, and if these are not correct he has a utility loss defined as the fraction of wealth he is willing to give up to know the "true" parameters. An MCMC estimation using Treasury yield data for the period 1971-2006 gives a posterior distribution of the "true" parameters. The posterior distribution gives a distribution of utility losses and the figure shows the distribution of these losses. The investor's relative risk aversion is $\gamma = 5$ while the investment horizon is $T = 5$.

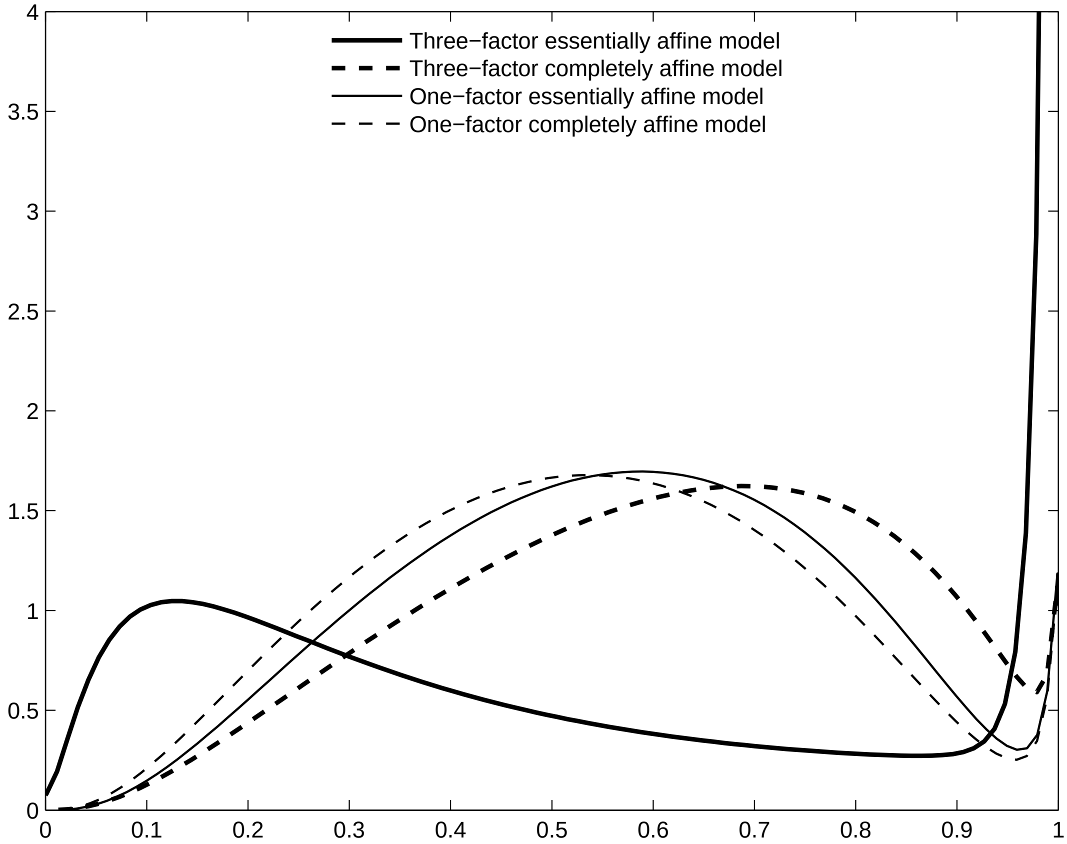

Figure 3: **Density of utility losses under parameter uncertainty and model misspecification.** It is assumed that the three-factor essentially affine model is the data-generating model. The investor chooses a portfolio strategy implied by his model parameter estimates, and if these are not correct he has a utility loss defined as the fraction of wealth he is willing to give up to know the "true" model and parameters. An MCMC estimation using Treasury yield data for the period 1971-2006 gives a posterior distribution of the "true" parameters of the three-factor essentially affine model. The posterior distribution gives a distribution of utility losses and the figure shows the distribution of these losses. The investor's relative risk aversion is $\gamma = 5$ while the investment horizon is $T = 5$.

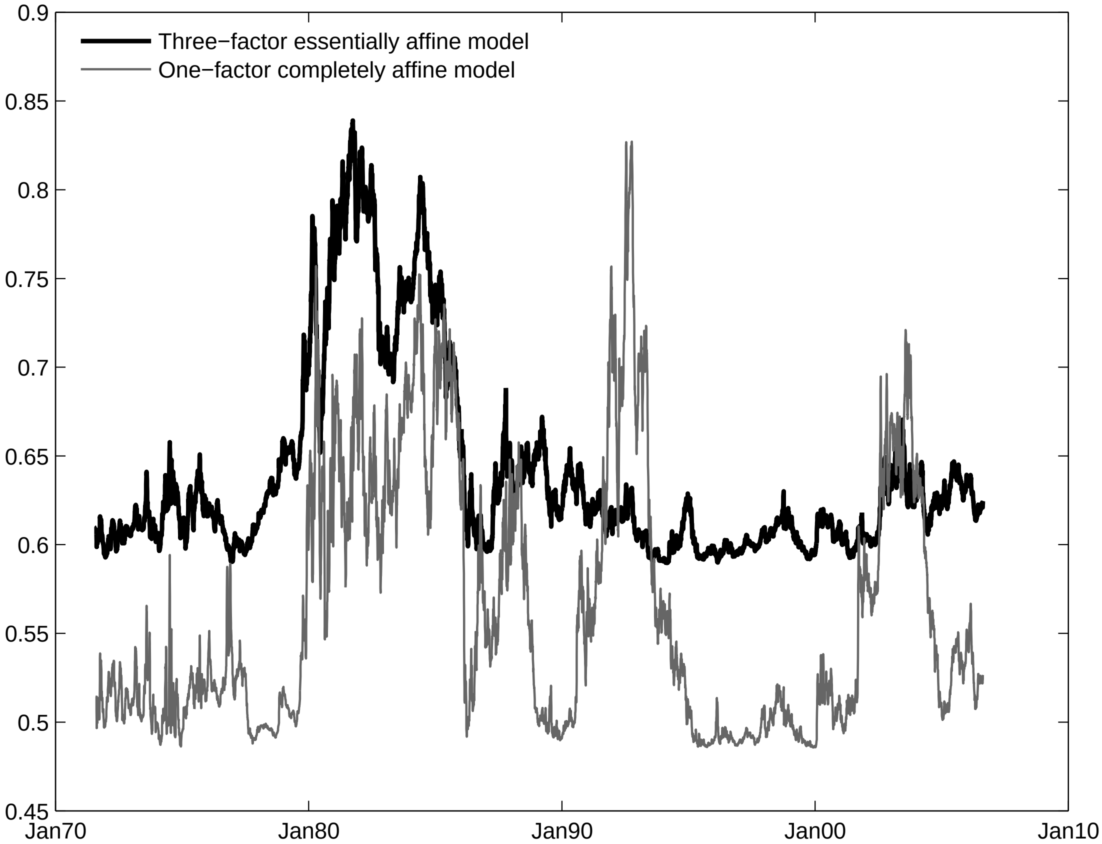

Figure 4: **Conditional expected utility losses over time when the three-factor essentially affine model is the data-generating model.** The figure shows daily conditional expected utility losses from August 16, 1971 to August 21, 2006 for an investor basing his investment on the three-factor essentially affine model and the one-factor completely affine model, respectively. The expected utility losses are conditional on the yield curve. The expected utility losses in the three-factor essentially affine model are due to parameter uncertainty, while in the one-factor completely affine model they are due to parameter uncertainty and model misspecification. The investor has an investment horizon of $T = 5$ years and a relative risk aversion of $\gamma = 5$.

δ_0		0.1589	
		(0.1582; 0.1629)	
	$i = 1$	$i = 2$	$i = 3$
δ_{Xi}	0.0027	0.0101	0.0101
	(0.0023; 0.0031)	(0.0099; 0.0106)	(0.0100; 0.0103)
κ_{1i}	1.5482	0.1433	0.0271
	(0.7776; 1.9290)	(−0.1706; 0.3661)	(−0.0760; 0.1824)
κ_{2i}	1.5041	0.4016	0.0709
	(0.5681; 1.7210)	(0.2052; 0.7384)	(−0.0693; 0.1990)
κ_{3i}	0.6401	−0.0272	0.1171
	(0.0553; 1.2182)	(−0.3259; 0.2367)	(0.0200; 0.2727)
λ_{0i}	0.1644	−0.4501	−1.081
	(−1.6999; 0.7431)	(−2.0522; 0.4381)	(−2.4085; −0.1181)
λ_{X1i}	−1.1519	−0.1433	−0.0271
	(−1.5380; 0.3850)	(−0.3670; 0.1702)	(−0.1829; 0.0760)
λ_{X2i}	−0.3433	0.4220	−0.0709
	(−0.5785; 0.5767)	(0.0853; 0.6191)	(−0.1995; 0.0692)
λ_{X3i}	−0.1961	0.4645	−0.0979
	(−0.7848; 0.3685)	(0.1990; 0.7546)	(−0.2558; −0.0011)

Table 1: **Parameter estimates for the three-factor essentially affine model.** The model is estimated using MCMC based on a panel data set of daily zero-coupon Treasury yields from 1971 to 2006. 95%-confidence intervals are given in parentheses.

δ_0		0.1826	
		(0.1809; 0.1838)	
	$i = 1$	$i = 2$	$i = 3$
δ_{Xi}	0.0027	0.0101	0.0104
	(0.0023; 0.0031)	(0.0098; 0.0104)	(0.0101; 0.0105)
κ_{1i}	0.3984	0	0
	(0.3925; 0.4055)	–	–
κ_{2i}	1.1272	0.8179	0
	(1.0928; 1.1620)	(0.8005; 0.8252)	–
κ_{3i}	0.4348	0.4367	0.0156
	(0.4259; 0.4695)	(0.4250; 0.4562)	(0.0153; 0.0158)
λ_i	−0.0711	−0.4697	−0.4490
	(−0.2974; 0.3602)	(−0.7459; −0.1164)	(−0.8023; −0.1372)

Table 2: **Parameter estimates for the three-factor completely affine model.** The model is estimated using MCMC based on a panel data set of daily zero-coupon Treasury yields from 1971 to 2006. 95%-confidence intervals are given in parentheses.

	Essentially affine model	Completely affine model
δ_0	-0.3869	-0.2011
	$(-0.3904; -0.3846)$	$(-0.2026; -0.1999)$
δ_X	0.0128	0.0055
	$(0.0127; 0.0128)$	$(0.0054; 0.0055)$
κ	0.1089	$2.72e-07$
	$(0.0089; 0.2573)$	$(1.13e-08; 1.63e-06)$
λ_0	3.9370	-0.0507
	$(-0.3345; 9.1777)$	$(-0.3664; 0.2729)$
λ_X	-0.1150	
	$(-0.2643; -0.0158)$	

Table 3: **Parameter estimates for the essentially affine and completely affine one-factor models.** The models are estimated using MCMC based on a panel data set of daily zero-coupon Treasury yields from 1971 to 2006. 95%-confidence intervals are given in parentheses.

Inv. horizon	π_{B1}	π_{B5}	π_{B10}	π_{rf}
	Panel A: Three-factor essentially affine model			
0	16.54	−2.34	0.37	−13.58
	(−40.16; 72.26)	(−8.03; 6.48)	(−1.74; 1.72)	(−63.53; 35.56)
5	12.92	−4.25	2.67	−10.34
	(−35.00; 64.70)	(−25.12; 14.05)	(−5.09; 10.74)	(−52.75, 25.15)
10	13.00	−5.18	3.78	−10.61
	(−35.37; 64.78)	(−25.94; 14.05)	(−5.27; 11.77)	(−53.01; 25.02)
	Panel B: Three-factor completely affine model			
0	18.01	−3.52	0.83	−14.33
	(5.56; 25.68)	(−6.91; 2.76)	(−1.82; 2.48)	(−21.01; −5.24)
5	18.01	−2.72	0.83	−15.13
	(5.56; 25.68)	(−6.11; 3.56)	(−1.82; 2.48)	(−21.81; −6.04)
10	18.01	−3.52	1.63	−15.13
	(5.56; 25.68)	(−6.91; 2.76)	(−1.02; 3.28)	(−21.81; −6.04)

Panel C: One-factor models

	Essentially affine		Completely affine	
Inv. horizon	π_{B5}	π_{rf}	π_{B5}	π_{rf}
0	0.28	0.72	0.37	0.63
	(−0.76; 2.03)	(−0.68; 1.36)	(−2.01; 2.66)	(−1.69; 3.00)
5	0.21	0.79	1.17	−0.17
	(−1.02; 1.90)	(−0.94; 2.01)	(−1.21; 3.46)	(−2.49; 2.20)
10	0.15	0.85	1.97	−0.97
	(−1.17; 1.84)	(−0.85; 2.17)	(−0.41; 4.26)	(−3.29; 1.40)

Table 4: **Portfolio weights for completely and essentially affine one- and three-factor models for an investor with a risk aversion of $\gamma = 5$.** Panel A shows the optimal portfolios for different horizons for an investor who bases his investment strategy on a three-factor essentially affine model. Panels B and C show the optimal portfolios for the three-factor completely affine model and the two one-factor models. In the one-factor models the investor invests only in the 5-year bond and the risk-free asset. The economy is assumed to be in steady state. An investment horizon of 0 corresponds to a myopic investment strategy and the portfolio weights are in natural units, not in percent. The estimates are based on an MCMC estimation using Treasury yield data for the period 1971-2006 and 95%-confidence intervals are given in parentheses.

Model	$T = 1$	$T = 5$	$T = 10$
	Panel A: $\gamma = 2$		
3-factor ess.	17.24%	59.28%	82.09%
3-factor com.	2.39%	11.07%	20.22%
1-factor ess.	4.59%	15.27%	27.93%
1-factor com.	0.71%	3.43%	6.54%
	Panel B: $\gamma = 5$		
3-factor ess.	11.36%	66.09%	87.00%
3-factor com.	0.97%	4.69%	9.03%
1-factor ess.	2.92%	11.31%	27.72%
1-factor com.	0.29%	1.41%	2.77%
	Panel C: $\gamma = 10$		
3-factor ess.	8.12%	70.09%	88.69%
3-factor com.	0.49%	2.39%	4.69%
1-factor ess.	2.11%	8.62%	26.82%
1-factor com.	0.14%	0.71%	1.41%

Table 5: **Expected utility loss due to parameter uncertainty.** For each of the four models in the table it is assumed that the model is the data-generating model and expected utility losses reflect the fraction of wealth an investor is willing to give up to avoid parameter uncertainty. Expected utility losses are calculated as explained in Section 4.3 and based on an MCMC estimation using Treasury yield data for the period 1971-2006. γ denotes the relative risk aversion of the investor while T is the investor's investment horizon in years.

Model	$T = 1$	$T = 5$	$T = 10$
Panel A: $\gamma = 2$			
3-factor ess.	0.01%	0.05%	0.10%
3-factor com.	3.7×10^{-3} %	0.02%	0.04%
1-factor ess.	1.4×10^{-5} %	1.2×10^{-4} %	3.4×10^{-4} %
1-factor com.	1.6×10^{-7} %	8.1×10^{-7} %	1.6×10^{-6} %
Panel B: $\gamma = 5$			
3-factor ess.	3.2×10^{-1} %	0.03%	0.07%
3-factor com.	1.5×10^{-3} %	7.4×10^{-3} %	0.01%
1-factor ess.	6.0×10^{-6} %	6.4×10^{-5} %	2.3×10^{-4} %
1-factor com.	6.4×10^{-8} %	3.2×10^{-7} %	6.4×10^{-7} %
Panel C: $\gamma = 10$			
3-factor ess.	1.7×10^{-3}%	0.02%	0.05%
3-factor com.	7.4×10^{-4} %	3.7×10^{-3} %	7.4×10^{-3} %
1-factor ess.	3.1×10^{-6} %	3.6×10^{-5} %	1.4×10^{-4} %
1-factor com.	3.2×10^{-8} %	1.6×10^{-7} %	3.2×10^{-7} %

Table 6: **Expected utility loss when market prices of risk parameters are known.** For each of the four models in the table it is assumed that the model is the data-generating model and that the market prices of risk parameters λ_0 and λ_X are known. Expected utility losses are calculated as explained in Section 4.3 and based on an MCMC estimation using Treasury yield data for the period 1971–2006. γ denotes the relative risk aversion of the investor, whereas T is the investor's investment horizon in years.

Model	$T=1$	$T=5$	$T=10$
Panel A: $\gamma=2$			
3-factor ess.	17.24%	59.28%	82.09%
3-factor com.	22.24%	81.67%	97.05%
1-factor ess.	26.25%	80.68%	96.21%
1-factor com.	25.24%	79.11%	95.44%
Panel B: $\gamma=5$			
3-factor ess.	11.36%	66.09%	87.00%
3-factor com.	11.77%	62.16%	88.95%
1-factor ess.	13.56%	57.16%	83.52%
1-factor com.	12.96%	54.16%	79.38%
Panel C: $\gamma=10$			
3-factor ess.	8.12%	70.09%	88.69%
3-factor com.	7.12%	44.22%	75.24%
1-factor ess.	8.20%	41.47%	71.42%
1-factor com.	7.78%	36.90%	61.26%

Table 7: **Expected utility loss due to parameter uncertainty and model misspecification.** It is assumed that the three-factor essentially affine model is the data-generating model. Expected utility losses reflect the fraction of wealth an investor is willing to give up to avoid parameter uncertainty and model misspecification. For the three-factor essentially affine model losses are only due to parameter uncertainty. Expected utility losses are calculated as explained in Section 4.3 and based on an MCMC estimation using Treasury yield data for the period 1971-2006. γ denotes the relative risk aversion of the investor while T is the investor's investment horizon in years.