Online Appendix to "Bond Return Predictability: Economic Value and Links to the Macroeconomy"

This online appendix is divided into four sections. In section A we perform pairwise tests aiming at disentangling more precisely the sources of the economic gains uncovered in Section 5 in the main body of the paper. In the first set of pairwise tests we compare the performance across model specifications (i.e., LIN, SV, TVP and TVPSV); in the second set of tests we compare across predictor variables (i.e., FB, CP, LN and FB+CP+LN). Section B computes out-of-sample R^2, predictive likelihood and CER values for the various model specifications relative to an EH benchmark augmented to incorporate stochastic volatility (EH-SV). In Section C, we quantify the out-of-sample economic gains using the Θ performance measure proposed by Ingersoll et al. (2007). Finally, in Section D, we relate our findings to Piazzesi et al. (2015).

Appendix A Pairwise Tests of Equality of Forecasting Performance

The results in Tables 3, 4 and 5 in the main text do not show that one modeling approach uniformly dominates the others. Moreover, the results do not show whether the out-of-sample performance values of the different model specifications (LIN, SV, TVP and TVPSV) are statistically different across models. To establish whether this is the case, we perform the following test. For each predictor variable (FB, CP, LN and FB + CP + LN) and each bond maturity (2, 3, 4, and 5 years) we run pairwise tests across the different modelling approaches. In particular, we test LIN against SV, LIN against TVP, LIN against TVPSV, SV against TVP, SV against TVPSV and finally TVP against TVPSV. The results are displayed in Table A-1 below. Panel A (B) displays CER values for an investor with mean variance (power) utility, while Panels C and D show values of the out-of-sample R^2 and predictive likelihood, respectively. Positive values suggest that the second model in the pair-wise comparison dominates the first model, while negative values suggest that the first model is best.

Starting from column (1), we find that the SV specification leads to substantial improvements over LIN –expect for the FB predictor– in both panels A and B. Slightly stronger results are obtained when comparing TVPSV and LIN in column (3). Conversely, column (2) shows that TVP does not systematically improve on LIN, and it is often worse than SV as shown by the fact that most of the values in column (4) are negative. Column (5) shows that the TVPSV specification is mostly statistically indistinguishable from SV. Finally, column (6) shows that the TVPSV specification leads to better performance than the TVP approach.

The results for the out-of-sample R^2 reported in Panel C suggest that this metric is less powerful in identifying differences between the model specifications' ability to generate accurate point forecasts.

The values of the predictive likelihood indicate that the differences in economic gains reported in Panels A and B are driven by the fact that the TVPSV and SV specifications capture the volatility dynamics in bond returns far better than the models with constant volatility. Indeed all values in columns (1), (3) and (5) of Panel D are positive and statistically significant.

Table A-1. **Pairwise Tests of Differences in Performance Across Model Specifications**

Panel A: CER, Mean Variance Utility

	(1)	(2)	(3)	(4)	(5)	(6)
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
FB 2y	-0.14%	0.22%***	0.42%**	0.36%**	0.56%**	0.20%*
FB 3y	0.12%	0.13%**	0.47%***	0.01%	0.35%**	0.34%**
FB 4y	0.28%	0.10%*	0.51%**	-0.17%	0.23%*	0.41%*
FB 5y	0.59%*	-0.01%	0.64%*	-0.60%	0.05%	0.66%**
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
CP 2y	0.37%**	0.11%**	0.47%**	-0.26%	0.10%	0.36%**
CP 3y	0.83%***	0.09%*	1.02%***	-0.74%	0.19%	0.93%***
CP 4y	0.72%***	0.06%	0.97%***	-0.66%	0.25%*	0.91%***
CP 5y	0.70%***	0.00%	0.68%**	-0.70%	-0.02%	0.67%**
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
LN 2y	-0.06%	0.00%	-0.01%	0.06%	0.05%	-0.01%
LN 3y	0.26%*	-0.04%	0.18%	-0.29%	-0.07%	0.22%
LN 4y	0.64%***	-0.02%	0.69%***	-0.65%	0.05%	0.70%***
LN 5y	0.91%***	0.02%	0.97%***	-0.89%	0.05%	0.95%***
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
$FB+CP+LN$ 2y	0.12%	0.07%	0.30%*	-0.05%	0.18%	0.23%*
$FB+CP+LN$ 3y	0.41%**	0.04%	0.63%***	-0.36%	0.23%*	0.59%***
$FB+CP+LN$ 4y	0.56%**	-0.01%	0.56%**	-0.58%	-0.01%	0.57%**
$FB+CP+LN$ 5y	0.99%***	-0.03%	0.91%**	-1.02%	-0.09%	0.93%**

Panel B: CER, Power Utility

	(1)	(2)	(3)	(4)	(5)	(6)
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
FB 2y	-0.18%	0.22%***	0.40%**	0.40%**	0.58%**	0.18%*
FB 3y	0.06%	0.13%**	0.43%**	0.07%	0.38%**	0.30%**
FB 4y	0.26%	0.12%*	0.50%**	-0.14%	0.24%*	0.38%*
FB 5y	0.55%*	-0.00%	0.59%*	-0.55%	0.04%	0.59%*
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
CP 2y	0.33%**	0.10%*	0.42%**	-0.23%	0.09%	0.33%**
CP 3y	0.79%***	0.09%*	0.98%***	-0.70%	0.18%	0.89%***
CP 4y	0.75%***	0.07%	1.00%***	-0.68%	0.25%*	0.93%***
CP 5y	0.66%***	0.01%	0.63%**	-0.66%	-0.04%	0.62%**
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
LN 2y	-0.07%	-0.00%	-0.00%	0.07%	0.07%	0.00%
LN 3y	0.23%	-0.03%	0.14%	-0.26%	-0.08%	0.18%
LN 4y	0.62%***	-0.01%	0.66%**	-0.63%	0.04%	0.67%***
LN 5y	0.92%***	0.03%	0.95%***	-0.89%	0.03%	0.92%***
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
$FB+CP+LN$ 2y	0.10%	0.07%	0.29%*	-0.03%	0.20%	0.22%*
$FB+CP+LN$ 3y	0.41%**	0.06%	0.63%***	-0.35%	0.23%*	0.57%***
$FB+CP+LN$ 4y	0.55%**	-0.01%	0.54%**	-0.56%	-0.00%	0.56%**
$FB+CP+LN$ 5y	0.96%***	-0.01%	0.86%**	-0.96%	-0.10%	0.86%**

Panel C: Out-of-sample R^2

	(1)	(2)	(3)	(4)	(5)	(6)
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
FB 2y	-0.38%	0.85%***	1.81%***	1.22%**	2.18%***	0.98%***
FB 3y	-0.00%	0.30%*	0.66%**	0.31%	0.66%**	0.36%
FB 4y	0.07%	0.14%	0.25%	0.07%	0.18%	0.11%
FB 5y	0.06%	0.01%	0.11%	-0.04%	0.05%	0.10%
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
CP 2y	0.25%*	0.40%**	1.42%***	0.15%	1.18%***	1.03%***
CP 3y	-0.06%	0.18%*	0.34%*	0.24%**	0.41%*	0.16%
CP 4y	-0.13%	-0.01%	0.14%	0.11%	0.27%*	0.15%
CP 5y	-0.11%	0.02%	-0.06%	0.13%	0.05%	-0.08%
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
LN 2y	1.74%***	0.48%**	2.40%***	-1.28%	0.68%**	1.94%***
LN 3y	0.13%	-0.11%	0.15%	-0.23%	0.02%	0.26%*
LN 4y	-0.13%	-0.21%	-0.22%	-0.08%	-0.09%	-0.01%
LN 5y	-0.15%	-0.03%	-0.27%	0.12%	-0.12%	-0.24%
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
$FB + CP + LN$ 2y	1.67%***	0.84%**	2.55%***	-0.84%	0.90%*	1.73%***
$FB + CP + LN$ 3y	0.35%	-0.25%	0.26%	-0.60%	-0.09%	0.50%*
$FB + CP + LN$ 4y	0.11%	-0.22%	-0.11%	-0.33%	-0.21%	0.11%
$FB + CP + LN$ 5y	0.07%	-0.24%	-0.19%	-0.31%	-0.27%	0.04%

Panel D: Predictive Likelihoods

	(1)	(2)	(3)	(4)	(5)	(6)
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
FB 2y	0.312***	0.004***	0.312***	-0.308	-0.000	0.308***
FB 3y	0.183***	0.003***	0.182***	-0.180	-0.001	0.180***
FB 4y	0.120***	0.001	0.119***	-0.119	-0.001	0.118***
FB 5y	0.085***	0.000	0.086***	-0.085	0.000	0.085***
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
CP 2y	0.302***	-0.001	0.295***	-0.303	-0.007	0.296***
CP 3y	0.179***	0.000	0.177***	-0.179	-0.002	0.176***
CP 4y	0.118***	-0.000	0.118***	-0.118	0.001	0.119***
CP 5y	0.085***	0.002***	0.085***	-0.083	-0.000	0.082***
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
LN 2y	0.293***	0.000	0.290***	-0.293	-0.004	0.289***
LN 3y	0.178***	0.000	0.176***	-0.177	-0.002	0.176***
LN 4y	0.119***	-0.001	0.117***	-0.119	-0.001	0.118***
LN 5y	0.084***	-0.000	0.084***	-0.084	-0.000	0.084***
	LIN vs SV	LIN vs TVP	LIN vs TVPSV	SV vs TVP	SV vs TVPSV	TVP vs TVPSV
$FB + CP + LN$ 2y	0.304***	0.004***	0.295***	-0.301	-0.009	0.291***
$FB + CP + LN$ 3y	0.181***	0.001*	0.176***	-0.179	-0.004	0.175***
$FB + CP + LN$ 4y	0.121***	0.002**	0.117***	-0.119	-0.004	0.115***
$FB + CP + LN$ 5y	0.085***	-0.000	0.084***	-0.086	-0.002	0.084***

This table displays the results of one-sided pairwise tests of differences in performance between the four models used in the paper (LIN, SV, TVP and TVPSV) across predictor variables (FB, CP, LN and FB+CP+LN) and bond maturities (2, 3, 4 and 5 years). Panels A and B report annualized CER values for an investor with power and mean-variance utility respectively, assuming a coefficient of relative risk aversion of five and weights on the bond positions constrained to lie between -1 and 2; Panel C shows out-of-sample R^2 values and Panel D shows values of the predictive likelihood. In each column, the null is that the two listed models have identical performance against the alternative that the second model is superior. Thus, in column (1) the null hypothesis is that the performance of the constant coefficients, constant volatility model (LIN) is the same as that of the model that allows for stochastic volatility (SV), while the alternative is that the latter is superior. Positive values suggest that the second model (SV) is better than the first model (LIN), while negative values suggest the reverse. A similar interpretation holds for the other pair-wise comparisons conducted in columns (2)-(6). P-values in Panels A and B are based on the Diebold-Mariano test while p-values in Panel C are based on the equal predictive accuracy test suggested by Clark and West (2007). Finally, to compute p-values in Panel D we follow Clark and Ravazzolo (2015) and apply the Diebold and Mariano (1995) t-test for equality of the average log-scores. The evaluation sample is 1990:01-2015:12. * significance at 10% level; ** significance at 5% level; *** significance at 1% level.

Next, we perform a set of model comparisons across the choice of predictor variables. The results in tables 3-5 in the main body of the paper suggest that the inclusion of the LN factor is important to our ability to generate out-of-sample statistical and economic gains. To establish more formally whether this is the case, we next perform the following test. For each model (LIN, SV, TVP and TVPSV) and each bond maturity (2, 3, 4, and 5 years) we run pairwise tests across different choices of the predictor variables. In particular, we test FB against CP, FB against LN, CP against LN, FB against FB+CP+LN, CP against FB+CP+LN and LN against FB+CP+LN. The results are displayed in Table A-2. Panel A (B) displays the CERs for an investor with mean variance (power) utility.

As highlighted in columns (2) and (3), LN generates higher economic gains than FB and CP. The CER values are significant in half of the cases compared to FB and are always significant (expect for the TVPSV and SV models for the 2-year bond) compared to CP. The positive values in columns (4) and (5) indicate that the trivariate model ($FB+CP+LN$) also leads to higher economic gains compared with the univariate specifications which include the FB or the CP factor, in all of the cases considered the CER values are significant at least at the 5% level. Finally, none of the CER values in column (6) are statistically significant and so the trivariate model does not seem to systematically improve over the LN factor, suggesting that the performance of the trivariate model is mainly driven by the LN factor.

Turning to the statistical performance measures, Panel C in the table below show strong evidence that, across bond maturities and model specifications, including the LN predictor leads to significantly higher R^2_{Oos} values compared to the models that exclude this variable. Hence, the LN factor leads to more accurate point forecasts. There is less evidence that this predictor matters to the predictive likelihood values which are more sensitive to how volatility dynamics is modeled.

Overall, we conclude from this new empirical evidence that the inclusion of the LN factor has an important role in uncovering both statistical (R^2_{Oos}) and economic gains (CER) from bond return predictability.

Table A-2. **Pairwise Tests of Differences in Performance Across Predictor Variables**

Panel A: CER, Mean Variance Utility

	(1)	(2)	(3)	(4)	(5)	(6)
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
LIN 2y	0.02%	0.70%**	0.69%***	0.65%**	0.64%***	-0.05%
LIN 3y	-0.50%	0.97%**	1.48%***	0.94%***	1.45%***	-0.03%
LIN 4y	-0.92%	0.65%	1.56%***	0.97%***	1.89%***	0.32%
LIN 5y	-0.99%	0.36%	1.35%**	0.78%**	1.76%***	0.41%
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
SV 2y	0.53%*	0.78%**	0.25%	0.91%***	0.39%*	0.13%
SV 3y	0.21%	1.11%**	0.90%**	1.23%***	1.02%***	0.12%
SV 4y	-0.47%	1.01%*	1.48%***	1.26%***	1.73%***	0.25%
SV 5y	-0.87%	0.68%	1.56%**	1.18%***	2.06%***	0.50%
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
TVP 2y	-0.10%	0.48%*	0.58%**	0.50%**	0.60%***	0.02%
TVP 3y	-0.55%	0.81%**	1.35%***	0.85%***	1.40%***	0.05%
TVP 4y	-0.96%	0.53%	1.49%***	0.85%**	1.81%***	0.33%
TVP 5y	-0.97%	0.39%	1.36%**	0.77%**	1.73%***	0.37%
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
$TVPSV$ 2y	0.06%	0.27%	0.21%	0.53%**	0.47%**	0.26%*
$TVPSV$ 3y	0.04%	0.69%*	0.64%**	1.11%***	1.06%***	0.42%*
$TVPSV$ 4y	-0.45%	0.82%*	1.28%***	1.02%***	1.47%***	0.19%
$TVPSV$ 5y	-0.95%	0.69%	1.64%***	1.04%**	1.99%***	0.36%

Panel B: CER, Power Utility

	(1)	(2)	(3)	(4)	(5)	(6)
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
LIN 2y	0.05%	0.70%**	0.65%**	0.67%**	0.62%***	-0.03%
LIN 3y	-0.48%	0.98%**	1.46%***	0.96%***	1.44%***	-0.02%
LIN 4y	-0.89%	0.72%*	1.61%***	1.04%***	1.92%***	0.32%
LIN 5y	-0.96%	0.39%	1.35%**	0.81%**	1.77%***	0.42%
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
SV 2y	0.56%*	0.81%**	0.25%	0.95%***	0.39%	0.14%
SV 3y	0.26%	1.15%**	0.89%**	1.32%***	1.06%***	0.16%
SV 4y	-0.40%	1.08%*	1.48%***	1.32%***	1.72%***	0.25%
SV 5y	-0.84%	0.76%	1.61%**	1.22%***	2.07%***	0.46%
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
TVP 2y	-0.08%	0.48%*	0.55%**	0.52%**	0.60%***	0.04%
TVP 3y	-0.52%	0.82%**	1.34%***	0.89%***	1.41%***	0.07%
TVP 4y	-0.94%	0.59%	1.53%***	0.91%**	1.84%***	0.31%
TVP 5y	-0.95%	0.43%	1.37%**	0.81%**	1.76%***	0.39%
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
$TVPSV$ 2y	0.07%	0.30%	0.23%	0.57%***	0.50%**	0.27%*
$TVPSV$ 3y	0.07%	0.69%*	0.63%**	1.16%***	1.10%***	0.47%*
$TVPSV$ 4y	-0.38%	0.88%*	1.27%***	1.08%***	1.47%***	0.20%
$TVPSV$ 5y	-0.92%	0.75%	1.67%***	1.08%**	2.00%***	0.33%

Panel C: Out-of-sample R^2

	(1)	(2)	(3)	(4)	(5)	(6)
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
LIN 2y	-0.68%	2.43%***	3.09%***	2.78%***	3.44%***	0.36%*
LIN 3y	-0.92%	2.77%**	3.65%***	3.30%***	4.18%***	0.54%*
LIN 4y	-1.04%	2.21%**	3.22%***	2.97%**	3.98%***	0.78%*
LIN 5y	-0.93%	1.83%**	2.73%***	2.65%**	3.55%***	0.84%*
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
SV 2y	-0.06%	4.48%***	4.54%***	4.77%***	4.82%***	0.29%
SV 3y	-0.98%	2.89%***	3.83%***	3.64%***	4.57%***	0.77%*
SV 4y	-1.24%	2.01%**	3.21%***	3.01%**	4.20%***	1.02%**
SV 5y	-1.10%	1.62%*	2.69%***	2.67%**	3.72%***	1.06%**
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
TVP 2y	-1.14%	2.07%***	3.17%***	2.78%***	3.87%***	0.73%*
TVP 3y	-1.04%	2.37%**	3.38%***	2.76%**	3.76%***	0.40%
TVP 4y	-1.19%	1.87%**	3.02%***	2.63%**	3.77%***	0.78%*
TVP 5y	-0.92%	1.79%**	2.68%***	2.41%**	3.30%***	0.63%*
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
$TVPSV$ 2y	-1.09%	3.02%***	4.06%***	3.52%***	4.55%***	0.51%*
$TVPSV$ 3y	-1.24%	2.27%**	3.46%***	2.90%**	4.09%***	0.65%*
$TVPSV$ 4y	-1.15%	1.75%**	2.87%***	2.63%**	3.74%***	0.90%*
$TVPSV$ 5y	-1.10%	1.46%*	2.53%***	2.36%**	3.42%***	0.91%*

Panel D: Predictive Likelihoods

	(1)	(2)	(3)	(4)	(5)	(6)
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
LIN 2y	0.001	0.004	0.003	0.007*	0.006*	0.003
LIN 3y	0.001	0.006*	0.005	0.009**	0.007**	0.002
LIN 4y	0.000	0.007*	0.007*	0.010**	0.010**	0.003*
LIN 5y	-0.002	0.006	0.007**	0.010**	0.011**	0.004*
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
SV 2y	-0.009	-0.014	-0.005	-0.000	0.009	0.014*
SV 3y	-0.003	0.001	0.004	0.006	0.009	0.005
SV 4y	-0.002	0.005	0.007	0.010*	0.013*	0.005
SV 5y	-0.002	0.005	0.007	0.010*	0.012*	0.005
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
TVP 2y	-0.004	0.001	0.004	0.007*	0.010***	0.006***
TVP 3y	-0.001	0.004	0.005	0.008*	0.008**	0.004**
TVP 4y	-0.001	0.005	0.006*	0.011**	0.012***	0.005**
TVP 5y	0.000	0.005	0.005	0.009*	0.009*	0.004**
	FB vs CP	FB vs LN	CP vs LN	FB vs FB,CP,LN	CP vs FB,CP,LN	LN vs FB,CP,LN
$TVPSV$ 2y	-0.016	-0.018	-0.002	-0.010	0.006	0.008*
$TVPSV$ 3y	-0.004	-0.000	0.004	0.003	0.007	0.003
$TVPSV$ 4y	-0.001	0.005	0.005	0.008	0.008	0.003
$TVPSV$ 5y	-0.003	0.004	0.007	0.008	0.010	0.004

This table displays the results of one-sided pairwise tests of differences in performance between the predictor variables used in the paper (FB, CP, LN, FB+CP+LN) across model specifications (LIN, SV, TVP and TVPSV) and bond maturities (2, 3, 4 and 5 years). Panels A and B report annualized CER values for an investor with power and mean-variance utility respectively, assuming a coefficient of relative risk aversion of five and weights on the bond positions constrained to lie between -1 and 2; Panel C shows out-of-sample R^2 values and Panel D shows values of the predictive likelihood. In each column, the null is that the two listed models have identical performance against the alternative that the second model is superior. Thus, in column (1) the null hypothesis is that the performance of FB is the same as that of CP, while the alternative is that the latter is superior. Positive values suggest that the second model (CP) is better than the first model (FB), while negative values suggest the reverse. A similar interpretation holds for the other pair-wise comparisons conducted in columns (2)-(6). P-values in Panels A and B are based on the Diebold-Mariano test while p-values in Panel C are based on the equal predictive accuracy test suggested by Clark and West (2007). Finally, to compute p-values in Panel D we follow Clark and Ravazzolo (2015) and apply the Diebold and Mariano (1995) t-test for equality of the average log-scores. The evaluation sample is 1990:01-2015:12. * significance at 10% level; ** significance at 5% level; *** significance at 1% level.

Appendix B Augmenting the Expectation Hypothesis Benchmark with Stochastic Volatility

In this section we compute out-of-sample R^2, predictive likelihood, and CER values for each model specification using as a benchmark the Expectation Hypothesis model augmented with stochastic volatility. This benchmark is more difficult to beat than the commonly used EH model with constant volatility.

The out-of-sample R^2 values displayed in Table B-1 show that replacing the EH benchmark with the EH-SV only leads to small changes in the out-of-sample R^2 values. In contrast, changing to the EH-SV benchmark has a much bigger effect on the predictive likelihood tests (Table B-2). For example, the EH-SV benchmark produces notably better predictive likelihood values than the LIN and TVP models which assume constant volatility. The new EH-SV benchmark continues to be dominated by the SV and TVPSV models which differ from the EH-SV benchmark by allowing for time variation in the conditional mean.

Turning to the economic utility measure (Table B-3), for three of four maturities the SV and TVPSV models produce significantly higher CER values than the EH-SV benchmark for the models that include LN as a predictor. We conclude, therefore, that the economic gains reported in the main body of the paper are robust to the choice of the benchmark.

Table B-1. **Out-of-sample forecasting performance relative to the EH-SV benchmark:** R^2 **values**

| | Panel A: 2 years | | | | | | Panel B: 3 years | | | | |
Model	OLS	LIN	SV	TVP	TVPSV	Model	OLS	LIN	SV	TVP	TVPSV
FB	1.10%**	1.67%*	1.30%**	2.50%***	**3.45%***	FB	2.48%***	2.08%***	2.08%***	2.38%***	**2.73%***
CP	-1.61%	0.99%**	1.24%**	1.39%**	**2.40%**	CP	-0.47%	1.18%**	1.12%**	1.36%**	**1.52%**
LN	-3.62%	4.06%***	5.72%***	4.51%***	**6.36%***	LN	0.96%***	4.79%***	4.91%***	4.69%***	**4.93%***
$FB+CP+LN$	-4.95%	4.40%***	6.00%***	5.21%***	**6.84%***	$FB+CP+LN$	-0.73%	5.31%***	**5.64%***	5.07%***	5.55%***

| | Panel C: 4 years | | | | | | Panel D: 5 years | | | | |
Model	OLS	LIN	SV	TVP	TVPSV	Model	OLS	LIN	SV	TVP	TVPSV
FB	**2.90%***	2.20%***	2.27%***	2.34%***	2.45%***	FB	**3.01%***	2.13%***	2.19%***	2.15%***	2.24%***
CP	0.11%*	1.18%**	1.06%**	1.17%**	**1.32%**	CP	0.55%*	1.22%**	1.11%**	**1.25%**	1.16%**
LN	2.77%***	**4.36%***	4.24%***	4.16%***	4.15%***	LN	3.68%***	**3.92%***	3.78%***	3.89%***	3.66%***
$FB+CP+LN$	0.99%***	5.11%***	**5.21%***	4.90%***	5.01%***	$FB+CP+LN$	1.82%***	4.73%***	**4.80%***	4.50%***	4.54%***

This table reports out-of-sample R^2 values for four prediction models based on the Fama-Bliss (FB), Cochrane-Piazzesi (CP), and Ludvigson-Ng (LN) predictors fitted to monthly bond excess returns, rx_{t+1}, measured relative to the one-month T-bill rate. The R^2_{OoS} is measured relative to the EH model augmented with stochastic volatility (EH-SV): $R^2_{OoS} = 1 - \frac{\sum_{\tau=\underline{t}-1}^{\bar{t}-1}(rx_{t+1} - \hat{r}_{xt+1|t})^2}{\sum_{\tau=\underline{t}-1}^{\bar{t}-1}(rx_{t+1} - \bar{r}_{xt+1|t})^2}$ where $\hat{r}_{xt+1|t}$ is the conditional mean of bond returns based on a regression of monthly excess returns on an intercept and lagged predictor variable(s), x_t: $rx_{t+1} = \mu + \beta' x_t + \varepsilon_{t+1}$. $\bar{r}_{t+1|t}$ is the forecast from the EH model (with stochastic volatility) which assumes that the βs are zero. We report results for five specifications: (i) ordinary least squares (OLS), (ii) a linear specification with constant coefficients and constant volatility (LIN), (iii) a model that allows for stochastic volatility (SV), (iv) a model that allows for time-varying coefficients (TVP) and (v) a model that allows for both time-varying coefficients and stochastic volatility ($TVPSV$). The out-of-sample period starts in January 1990 and ends in December 2015. We measure statistical significance relative to the expectation hypothesis model using the Clark and West (2007) test statistic. * significance at 10% level; ** significance at 5% level; *** significance at 1% level. For every model and maturity, we denote in bold font the R^2_{OoS} of the estimation method (LIN, SV, TVP and TVPSV) which delivers the best result.

Table B-2. **Out-of-sample forecasting performance relative to the EH-SV benchmark: predictive likelihood**

Model	Panel A: 2 years				Panel B: 3 years			
	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	-0.293	**0.019****	-0.289	0.019***	-0.170	**0.013****	-0.168	0.012**
CP	-0.292	**0.010***	-0.293	0.003	-0.169	**0.010****	-0.169	0.008**
LN	-0.289	**0.004**	-0.288	0.001	-0.164	**0.014***	-0.164	0.012*
$FB + CP + LN$	-0.286	**0.018***	-0.282	0.009	-0.162	**0.019****	-0.160	0.015*

Model	Panel C: 4 years				Panel D: 5 years			
	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	-0.107	**0.013****	-0.107	0.012**	-0.076	0.009**	-0.075	**0.010****
CP	-0.107	0.010***	-0.108	**0.011****	-0.077	**0.008****	-0.075	0.007**
LN	-0.101	**0.018****	-0.101	0.016**	-0.070	**0.014****	-0.070	0.014*
$FB + CP + LN$	-0.098	**0.023****	-0.096	0.019**	-0.066	**0.019****	-0.066	0.017*

This table reports the log predictive score for four forecasting models that allow for time-varying predictors relative to the log-predictive score computed under the expectation hypothesis model augmented with stochastic volatility (EH-SV). The four forecasting models use the Fama-Bliss (FB) forward spread predictor, the Cochrane-Piazzesi (CP) combination of forward rates, the Ludvigson-Ng (LN) macro factor, and the combination of these. Positive values of the test statistic indicate that the model with time-varying predictors generates more precise forecasts than the EH (with stochastic volatility) benchmark. We report results for a linear specification with constant coefficients and constant volatility (LIN), a model that allows for stochastic volatility (SV), a model that allows for time-varying coefficients (TVP) and a model that allows for both time-varying coefficients and stochastic volatility ($TVPSV$). The results are based on out-of-sample estimates over the sample period 1990 - 2015. ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. For every model and maturity, we denote in bold font the Predictive Likelihood of the estimation method (LIN, SV, TVP and TVPSV) which delivers the best result.

Table B-3. **Out-of-sample economic performance of bond portfolios relative to EH-SV benchmark**

Panel A: Power Utility

	Panel A.1: 2 years				Panel A.2: 3 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	-0.52%	-0.70%	-0.30%	**-0.12%**	-0.43%	-0.38%	-0.30%	**0.00%**
CP	-0.47%	-0.14%	-0.38%	**-0.05%**	-0.91%	-0.12%	-0.82%	**0.07%**
LN	**0.18%**	0.11%	0.18%	0.18%	0.55%	**0.78%****	0.52%	0.70%**
$FB + CP + LN$	0.15%	0.25%	0.22%	**0.44%****	0.53%	0.94%**	0.59%	**1.17%*****

	Panel A.3: 4 years				Panel A.4: 5 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.29%	0.55%	0.41%	**0.79%***	0.86%*	1.41%*	0.86%*	**1.45%****
CP	-0.60%	0.15%	-0.53%	**0.40%**	-0.10%	**0.57%**	-0.09%	0.53%
LN	1.01%**	1.63%***	1.00%**	**1.67%*****	1.25%**	2.17%***	1.28%**	**2.20%*****
$FB + CP + LN$	1.33%**	**1.87%*****	1.31%**	1.87%***	1.68%**	**2.63%*****	1.67%**	2.53%***

Panel B: Mean Variance Utility

	Panel B.1: 2 years				Panel B.2: 3 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	-0.54%	-0.68%	-0.31%	**-0.12%**	-0.40%	-0.28%	-0.27%	**0.07%**
CP	-0.52%	-0.15%	-0.41%	**-0.05%**	-0.90%	-0.07%	-0.82%	**0.12%**
LN	0.16%	0.11%	**0.17%**	0.16%	0.57%*	**0.83%****	0.54%*	0.76%**
$FB + CP + LN$	0.12%	0.24%	0.18%	**0.42%****	0.54%	0.95%**	0.59%	**1.18%*****

	Panel B.3: 4 years				Panel B.4: 5 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.37%	0.65%	0.47%	**0.88%***	0.86%*	1.45%**	0.85%*	**1.50%****
CP	-0.55%	0.18%	-0.48%	**0.43%**	-0.12%	**0.58%**	-0.12%	0.55%
LN	1.02%**	1.66%***	1.00%**	**1.70%*****	1.22%**	2.13%***	1.24%**	**2.19%*****
$FB + CP + LN$	1.34%**	**1.90%*****	1.33%**	1.90%***	1.64%**	**2.63%*****	1.61%**	2.55%***

This table reports annualized certainty equivalent return values for portfolio decisions based on recursive out-of-sample forecasts of bond excess returns. All values are measured relative to the benchmark of an expectations hypothesis model augmented with stochastic volatility (EH-SV). Each period an investor with power utility (Panel A) / mean-variance utility (Panel B) and coefficient of relative risk aversion of 5 selects 2, 3, 4, or 5-year bond and 1-month T-bills based on the predictive density implied by a given model. The four forecasting models use the Fama-Bliss (FB) forward spread predictor, the Cochrane-Piazzesi (CP) combination of forward rates, the Ludvigson-Ng (LN) macro factor, and the combination of these. We report results for a linear specification with constant coefficients and constant volatility (LIN), a model that allows for stochastic volatility (SV), a model that allows for time-varying coefficients (TVP) and a model with both time varying coefficients and stochastic volatility ($TVPSV$). Statistical significance is based on a one-sided Diebold-Mariano test applied to the out-of-sample period 1990-2015. * significance at 10% level; ** significance at 5% level; *** significance at 1% level. For every model and maturity, we denote in bold font the CER of estimation method (LIN, SV, TVP and TVPSV) which delivers the best result.

Appendix C Ingersoll et al. (2007) Performance Measure.

Ingersoll et al. (2007) establish a set of conditions under which the following Θ performance measure is manipulation-proof:

$$\Theta = \frac{12}{(1-A)} ln \Big[\frac{1}{T} \sum_{t=1}^{T} \Big(\frac{1+r_t}{1+r_{f,t}} \Big)^{1-A} \Big].$$

Here A denotes the investor's relative risk aversion, T denotes the length of the evaluation window, $r_{f,t}$ denotes the risk-free rate, and r_t denotes the realized net portfolio return of a given investment strategy. One additional benefit of this measure is that it alleviates concerns related to non-normality of the realized returns. Unfortunately no formal statistical test is available to assess whether the sample estimate of Θ is statistically different from zero. We therefore report CER values in the paper (see footnote 20 for more details on how we evaluate the statistical significance of the CERs).

We follow Thornton and Valente (2012) and Sarno et al. (2016) and replace $r_{f,t}$ with $r_{bench,t}$, the out-of-sample realized net portfolio return obtained under the Expectation Hypothesis benchmark and r_t with $r_{model,t}$, the out-of-sample realized net portfolio return under the alternative models in order to quantify the economic gains that these models generate in excess of the benchmark.

Results for the Θ performance measure are reported in Table C-1. Compared to the CER values reported in the paper, a very similar pattern emerges. First, the LN factor still delivers considerably better economic performance than the CP and FB factors. Second, we still find that in most of the cases the TVPSV model performs best. Finally, the economic gains tend to be larger for the longest bond maturities.

The fact that the CER and Θ values lead to similar conclusions is not surprising. As highlighted in Ingersoll et al. (2007) Θ "can be interpreted as the annualized continuously compounded excess certainly equivalent of the portfolio" and it "looks like the average of a power utility function, calculated over the return history".

Table C-1. **Out-of-sample economic performance of bond portfolios**

Panel A: Power Utility

	Panel A.1: 2 years				Panel A.2: 3 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	-0.49%	-0.69%	-0.26%	**-0.08%**	0.13%	0.18%	0.27%	**0.60%**
CP	-0.46%	-0.08%	-0.35%	**0.01%**	-0.45%	0.46%	-0.34%	**0.69%**
LN	0.18%	0.13%	0.18%	**0.20%**	1.05%	**1.33%**	1.02%	1.27%
$FB+CP+LN$	0.12%	0.24%	0.19%	**0.44%**	1.00%	1.45%	1.07%	**1.69%**

	Panel A.3: 4 years				Panel A.4: 5 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.91%	1.22%	1.03%	**1.49%**	1.38%	2.08%	1.39%	**2.16%**
CP	-0.19%	0.68%	-0.11%	**1.00%**	0.17%	0.94%	0.19%	**0.99%**
LN	1.54%	2.23%	1.53%	**2.29%**	1.71%	2.69%	1.74%	**2.76%**
$FB+CP+LN$	1.85%	2.48%	1.84%	**2.50%**	2.17%	**3.24%**	2.17%	3.18%

Panel B: Mean Variance Utility

	Panel B.1: 2 years				Panel B.2: 3 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	-0.50%	-0.66%	-0.26%	**-0.04%**	0.22%	0.35%	0.37%	**0.76%**
CP	-0.50%	-0.06%	-0.38%	**0.04%**	-0.42%	0.58%	-0.31%	**0.82%**
LN	0.21%	0.17%	0.21%	**0.23%**	1.19%	**1.52%**	1.16%	1.47%
$FB+CP+LN$	0.13%	0.28%	0.20%	**0.47%**	1.13%	1.61%	1.18%	**1.87%**

	Panel B.3: 4 years				Panel B.4: 5 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	1.02%	1.39%	1.14%	**1.66%**	1.42%	2.19%	1.42%	**2.28%**
CP	-0.13%	0.75%	-0.05%	**1.08%**	0.17%	1.00%	0.19%	**1.05%**
LN	1.65%	2.40%	1.63%	**2.47%**	1.77%	2.80%	1.79%	**2.89%**
$FB+CP+LN$	1.98%	2.67%	1.98%	**2.69%**	2.24%	**3.40%**	2.22%	3.34%

This table reports the annualized performance measure of Ingersoll et al. (2007) for portfolio decisions based on recursive out-of-sample forecasts of bond excess returns. Specifically, we compute $\frac{12}{(1-A)}ln\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1+r_{model,t}}{1+r_{bench,t}}\right)^{1-A}\right]$ where A denotes relative risk-aversion, $r_{bench,t}$ denotes the out-of-sample realized net portfolio return under the Expectation Hypothesis benchmark, and $r_{model,t}$ denotes the out-of-sample realized net portfolio return under the alternative models. Each period an investor with power utility (Panel A) / mean-variance utility (Panel B) and coefficient of relative risk aversion of 5 selects 2, 3, 4, or 5-year bond and 1-month T-bills based on the predictive density implied by a given model. The four forecasting models use the Fama-Bliss (FB) forward spread predictor, the Cochrane-Piazzesi (CP) combination of forward rates, the Ludvigson-Ng (LN) macro factor, and the combination of these. We report results for a linear specification with constant coefficients and constant volatility (LIN), a model that allows for stochastic volatility (SV), a model that allows for time-varying coefficients (TVP) and a model with both time varying coefficients and stochastic volatility ($TVPSV$). For every model and maturity, we denote in bold font the Θ of the estimation method (LIN, SV, TVP and TVPSV) which delivers the best result.

The table below summarizes the differences between Thornton and Valente (2012) , Sarno et al. (2016), and this paper.

Table C-2. **Comparison between Thornton and Valente (2012), Sarno et. al. (2016) and this paper.**

	Thornton et. al. (2012)	Sarno et. al. (2016)	This Paper
Asset Allocation	Multivariate	Univariate and Multivariate	Univariate and Multivariate
Utility Function	Mean-variance	Power and Mean-variance	Power and Mean-variance
Risk-Aversion	5	3	5
Performance Measure	Θ and Sharpe Ratio	Θ	Θ and CER
Lower Bound Constraint	-100%	-100%	-100%
Upper Bound Constraint	200%	200%	200%
Predictors	FB and CP	Not Applicable	FB, CP and LN
Bond Maturity	2, 3, 4 and 5 years	1 and 3 months; 1, 2 and 3 years	2, 3, 4 and 5 years

Appendix D   Economic Gains and difference between statistical and subjective interest rates

Using survey data on interest rate forecasts, Piazzesi et al. (2015) find that subjective risk premia are less volatile and less cyclical than statistical risk premia. The reason for the discrepancy is that survey forecasts of interest rates are made as if both the level and the slope of the yield curve are more persistent than under common statistical models.

Piazzesi et al. (2015) derive the following equation to construct subjective bond risk premia from survey data on interest rate forecasts:

$$E_t\left[rx_{t,t+h}^{(n)}\right] = E_t^*\left[rx_{t,t+h}^{(n)}\right] + (n-h)\left(E_t^*\left[i_{t+h}^{(n-h)}\right] - E_t\left[i_{t+h}^{(n-h)}\right]\right), \tag{A-1}$$

where $E_t\left[rx_{t,t+h}^{(n)}\right]$, the statistical premium, and $E_t\left[i_{t+h}^{(n-h)}\right]$, the statistical interest-rate expectation, are obtained from a VAR(1), and $E_t^*\left[i_{t+h}^{(n-h)}\right]$, the subjective interest-rate expectation, is obtained from the Blue Chip data.

To see whether the utility gains from our portfolio analysis might be related to biases in market participants' forecasts of future interest rates, we regress utility gains, computed relative to the EH benchmark, on the absolute difference between the subjective and the statistical interest rate forecasts, $|E_t^*[i_{t+h}^{(n-h)}] - E_t[i_{t+h}^{(n-h)}]|$.[1] Results from these regressions, reported in Table D-1, show a mostly positive correlation between utility gains and differences in the subjective and statistical interest rate forecasts.

---

[1]We also tried using the squared difference, $\left(E_t^*[i_{t+h}^{(n-h)}] - E_t[i_{t+h}^{(n-h)}]\right)^2$ and found similar results.

Table D-1. **Economic Gains and difference between statistical and subjective interest rates.**

| | Utility Gains | | | | | | | |
| | Power Utility | | | | Mean Variance Utility | | | |
	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.131**	0.223**	0.122**	0.213**	0.125*	0.260**	0.120*	0.258**
CP	-0.074	0.080	-0.027	0.117	-0.079	0.084	-0.031	0.134
LN	0.097	0.167*	0.094	0.174*	0.094	0.199*	0.097	0.208*
FB+CP+LN	0.160**	0.188*	0.189**	0.194*	0.167**	0.222**	0.202**	0.231**

This table displays the slope coefficient from regressing utility gains (with respect to the EH benchmark) on the absolute difference between the subjective and the statistical forecasts of interest rates. The subjective interest rate forecasts are based on the Blue Chip survey while the statistical interest rate forecasts are based on a VAR(1). The four forecasting models use the Fama-Bliss (FB) forward spread predictor, the Cochrane-Piazzesi (CP) combination of forward rates, the Ludvigson-Ng (LN) macro factor, and the combination of these. We report results for a linear specification with constant coefficients and constant volatility (LIN), a model that allows for stochastic volatility (SV), a model that allows for time-varying coefficients (TVP) and a model that allows for both time-varying coefficients and stochastic volatility ($TVPSV$). The results are based on out-of-sample estimates over the sample period 1990 - 2015 and use the two-year bond maturity. ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

References

Ingersoll, J., M. Spiegel, W. Goetzmann, and I. Welch (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies 20*(5), 1503–1546.

Piazzesi, M., J. Salomao, and M. Schneider (2015, March). Trend and cycle in bond premia. *Working Paper*.

Sarno, L., P. Schneider, and C. Wagner (2016). The economic value of predicting bond risk premia. *Journal of Empirical Finance 37*, 247 – 267.

Thornton, D. L. and G. Valente (2012). Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. *Review of Financial Studies 25*(10), 3141–3168.