BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2014-13

Bond Risk Premia and Gaussian Term Structure Models

by Bruno Feunou and Jean-Sébastien Fontaine

Bank of Canada Working Paper 2014-13

April 2014

Bond Risk Premia and Gaussian Term Structure Models

by

Bruno Feunou and Jean-Sébastien Fontaine

Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
bfeunou@bankofcanada.ca
jsfontaine@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

We thank Greg Bauer, Antonio Diez de los Rios, Greg Duffee, Anh Le, Dora Xia, Glen Keenleyside and participants at the 2013 Bank of Canada Fixed Income conference for comments and suggestions.

Abstract

Cochrane and Piazzesi (2005) show that (i) lagged forward rates improve the predictability of annual bond returns, adding to current forward rates, and that (ii) a Markovian model for monthly forward rates cannot generate the pattern of predictability in annual returns. These results stand as a challenge to modern Markovian dynamic term structure models (DTSMs). We develop the family of conditional mean DTSMs where the yield dynamics depend on current yields *and their history*. Empirically, we find that (i) current and past yields generate cyclical risk-premium variations, (ii) the model risk premia offer better returns forecasts, and (iii) the model coefficients are close to Cochrane-Piazzesi regressions of long-horizon returns. Yield decompositions differ significantly from what a standard model suggests – the expectation component decreases less in a recession and increases less in the recovery. A small Markovian factor "hidden" in measurement error (Duffee, 2011) explains some of the differences but is not sufficient to match the evidence.

JEL classification: E43, E47, G12
Bank classification: Interest rates; Asset pricing

Résumé

Cochrane et Piazzesi (2005) montrent que les taux à terme retardés améliorent la prévisibilité des rendements obligataires en complétant l'apport des taux à terme courants, et qu'un modèle markovien des taux à terme ne peut produire le profil de prévisibilité des rendements annuels. Ces résultats compliquent la tâche des modèles dynamiques de la structure par terme (DTSM) de type markovien. Nous construisons des DTSM nouveaux, où la dynamique de la moyenne conditionnelle dépend des taux courants *et passés*. De manière empirique, nous constatons que les taux passés contribuent aux variations cycliques de la prime de risque, que le modèle offre de meilleures prévisions des rendements et que ses coefficients avoisinent les résultats des régressions effectuées sur les rendements de long terme, prolongeant l'étude de Cochrane et Piazzesi. Comparativement au modèle standard, la décomposition des taux diffère sensiblement : la composante des anticipations diminue moins durant une récession et s'accroît moins en période de reprise. La présence dans les taux d'un facteur de risque markovien « caché » dans les erreurs de mesure ne peut rendre compte entièrement du phénomène.

Classification JEL : E43, E47, G12
Classification de la Banque : Taux d'intérêt; Évaluation des actifs

Introduction

This paper has its sources at the confluence of two stylized facts. On the one hand, there is a broad consensus that three factors summarize the cross-section of interest rates (Litterman and Scheinkman, 1991). On the other hand, we also know that current yields do not summarize predictable variations in future interest rates. For instance, Cochrane and Piazzesi (2005) devote one section to "the failure of Markovian models,"[1] documenting that past forward rates help forecast bond returns. The tension is particularly acute in dynamic term structure models (DTSMs) that use yields as risk factors. This paper offers a constructive resolution. We introduce the family of Conditional Mean DTSMs (CM-DTSM), capturing the cross-section of yields with three risk factors *and* matching the predictive content of lagged forward rates with no additional sources of risk. Our extension is parsimonious and intuitive, and estimation is straigthforward. It is also grounded in empirical evidence and the results have practical implications for researchers and investors alike: the implied risk-premium variations are sizably larger and more cyclical than in standard specifications.

To understand the underlying tension, it is useful to go back to the basic building blocks of Gaussian DTSMs (GDTSM). Consider a standard (exponential-affine) pricing kernel with $\mathcal{N}$ sources of risk. This is a statement about the risk-neutral distribution $\mathbb{Q}$: the cross-section of $\mathcal{J}$ yield portfolios $\mathcal{P}_t^o$ is linked to $\mathcal{N} < \mathcal{J}$ risk factors Z_t via a linear relationship. Next, consider the standard assumption that risk factors are Markovian. This is a statement about time series: conditional forecasts of $\mathcal{P}_{t+h}^o$ are linked to Z_t, via a linear relationship (for all horizons $h > 0$). Taken together, these assumptions suggest that current yields are all we need to construct our best forecasts of future yields. This proposition is rejected in the data.

This discussion also suggests that there is no necessary connection between (i) the number of shocks required to match the cross-section of yields and (ii) the information set $\mathcal{I}_t$ required for the conditional expectations $E[\mathcal{P}_{t+h}^o|\mathcal{I}_t]$. Regarding point (i), we make clear in the following that if yields are linear in $\mathcal{N}$ stationary Gaussian risk factors Z_t, then the no-arbitrage assumption requires that Z_t have Markovian dynamics under the $\mathbb{Q}$ measure. This result corresponds to the heuristic argument that today's prices should reflect all available information and should not generate profitable investment strategies unless associated with proportionate risks. We maintain this standard assumption. Regarding point (ii), and as a corollary, we conclude

[1] See their Section III, particularly their Section III.C, p. 153. In their words: "The importance of extra monthly lags means that a VAR(1) *monthly* representation, of the type specified by nearly every explicit term structure model, does not capture the patterns we see in *annual* return forecasts."

that we need to break away from $\mathcal{I}_t = \{\mathcal{P}^o_t\}$ to capture the predictive content of lagged forward rates.

The first part of the paper studies the empirical properties of $E[\mathcal{P}^o_{t+h}|\mathcal{I}_{\mathcal{P},t}]$, where we take $\mathcal{I}_{\mathcal{P},t} = \{\mathcal{P}^o_t, \mathcal{P}^o_{t-1}, \ldots\}$ the history of $\mathcal{P}^o_t$. We revisit the predictability of bond returns for horizons between one and twelve months, extending the results in Cochrane and Piazzesi (2005) for annual returns. First, we confirm that forward rates forecast bond returns at every horizon. Perhaps unsurprisingly, one combination of forward rates is enough in each case. Second, and this is more important, we find that lags of the return-forecasting factor add to the predictability evidence. In fact, the relative contribution of past forward rates increases as we shorten the returns horizon and peaks at the monthly horizon. This result stands in direct contrast with standard Markovian models. Third, the coefficients on additional lags of the returns-forecasting factor exhibit a clear decaying pattern. We show analytically and empirically that the evidence is consistent with models where the conditional mean $E[\mathcal{P}^o_{t+1}|\mathcal{I}_{\mathcal{P},t}]$ is Markovian but where $\mathcal{P}^o_t$ is not.

The second part of the paper introduces the class of CM-DTSMs. Consistent with the cross-section evidence, we specify that the $\mathcal{N}$ risk factors Z_t are Markovian under $\mathbb{Q}$. Consistent with the predictability evidence, we specify the process for Z_t under $\mathbb{P}$ via the dynamics of its conditional mean $\mathcal{E}_{Z,t} \equiv E[Z_{t+1}|\mathcal{I}_{Z,t}]$, with $\mathcal{I}_{Z,t} = \{Z_t, Z_{t-1}, \ldots\}$. To get a sense of its properties, this process is analogous to GARCH dynamics where the conditional variance is driven by the (squared) innovations in the underlying risk factors. Here, the conditional mean is driven by a rotation of the $\mathcal{N}$ innovations $u_t \equiv Z_t - \mathcal{E}_{Z,t}$. Conditional DTSMs have the following important property: there is a separation between the *spanning* role of Z_t and the *forecasting* role of $\mathcal{E}_{Z,t}$. The $\mathcal{N} \times \mathcal{N}$ parameter matrix θ controls this separation.[2] Indeed, if $\theta = 0$, all yields and all conditional forecasts are a function of Z_t, nesting the standard VAR(1) case (e.g., $\mathcal{E}_{Z,t} = \Phi Z_t$).

Finally, the empirical section makes two broad points. First, $\theta = 0$ is soundly rejected in a specification with five bond yields and five risk factors. This special case allows the risk factors to capture *all* the information from current yields and gives more chances to the standard model. Yet, the variability of the risk premium is often halved in models with $\theta = 0$ relative to a case where we account for the information

[2] The representation in terms of the conditional mean was introduced by Fiorentini and Sentana (1998) in the context of time-series models. The specific case we consider corresponds to VARMA(1,1) dynamics in the usual representation and nests the standard VAR(1) specification when $\theta = 0$.

in past yields. In addition, the risk premium offers much better forecasts of excess returns for those same bonds used for estimations of the model. In fact, a CM-DTSM model provides a close match to the forward rate coefficients estimated from Cochrane-Piazzesi regressions. In contrast, we find that predictive R^2s are halved and that the model does not reproduce the well-known tent shape when we impose $\theta = 0$. The deteriorations are much worse for shorter returns horizons. Second, we ask whether we can match both cross-section evidence and the predictability evidence with $\mathcal{N} = 3$ factors. The answer is yes. Three factors are sufficient to generate the cross-section evidence by construction. In addition, we find that reducing the number of shocks leads to no loss in the variability of the risk premium or in the predictability of bond excess returns. However, reducing the number of shocks widens the gap between the standard case and our CM-DTSM specification. We compare yield decompositions from different models: the risk premium is more cyclical in CM-DTSM models, and the expectation component plays a much lesser role in explaining yield changes in recessions.

We devote considerable space to comparing our approach with closely related results in Duffee (2011), where the combination of a 5-factor (Markovian) model and Kalman filtering produces a "hidden" risk factor. This hidden factor does not affect current yields, but it loads on past yields and generates substantial risk-premium variability. Duffee (2011) "advocates a significant change in the construction and estimation of multifactor term structure models." We agree. In fact, we argue that his findings are nested in our framework. To see the connection, consider the general case where every yield portfolio $\mathcal{P}_t^o$ is measured with errors. The information set available to the econometrician is $\mathcal{I}_{\mathcal{P},t}$. Once we condition on $\mathcal{I}_{\mathcal{P},t}$ (and not $\mathcal{I}_{Z,t}$), our main theorem shows that we can construct an equivalent steady-state CM-DTSM where the conditional mean $\mathcal{E}_{\mathcal{P},t} \equiv E[\mathcal{P}_t^o|\mathcal{I}_{\mathcal{P},t}]$ is Markovian, summarizing the dynamics of all observed portfolios $\mathcal{P}_t^o$, but with cross-equation restrictions due to the factor structure embedded in yields.

We identify two mechanisms that may coexist and generate a role for past yields within our maximally canonical form. The first mechanism corresponds to the effect of filtering through mismeasured yields, as in Duffee (2011). This mechanism exists even if Z_t is Markovian, in which case the observed yields have restricted CM dynamics. The restriction arises because of the additional structure in the covariance matrix of the measurement errors, R. The second mechanism arises when the risk factors themselves have CM dynamics. This mechanism exists even if yields are observed without errors, in which case the information content of past yields is parameterized by θ. Each mechanism is separately identified when $\mathcal{J} > \mathcal{N}$ portfolios are included in the model.

Since the two mechanisms may coexist, it is natural to ask whether measurement errors alone can rationalize the predictability evidence. We estimate several versions of the 3-factor model to answer this question. We consider cases where $R = \sigma^2 I$, or where R is a free diagonal covariance matrix, and cases with $\theta = 0$, or with θ unrestricted. The results are consistent with those in Duffee (2011). Filtering by itself adds considerable variability in the risk premium and improves the predictability of the risk premium. But the R^2s stand somewhere halfway between the standard case and our unrestricted CM-DTSM model. Hence we conclude that non-Markovian effects in the risk factors themselves are required to match the evidence. We also differ with Duffee (2011), since our results show that it is unnecessary to add more shocks, both in theory and empirically. Our 3-factor model matches the evidence almost as well as the best 5-factor model, but it is considerably more parsimonious. In addition, there is no need to appeal to a special structure in the risk premium to hide the additional factors in tiny measurement errors. Instead, breaking away from the Markovian assumption introduces separate spanning and forecasting roles for $\mathcal{P}_t^o$ and $\mathcal{E}_{\mathcal{P},t}$, with no additional shocks.[3]

Our approach is equivalent to generalizing the prices of risk to an affine function of Z_t and $\mathcal{E}_{Z,t-1}$. This is consistent with existing equilibrium models where, for instance, agents are learning about the underlying process, or because of adaptive expectations. Froot (1989) points out that the usual tests of the expectations hypothesis rely on the maintained assumption that investors' expectations are rational, and argues that expectational errors play a significant role. Piazzesi and Schneider (2009) provide additional evidence that subjective expected excess returns are less volatile and less cyclical. Recently, Cieslak and Povala (2013) have also recognized the significance of lagged information and explored the role of informational frictions related to agents' perceptions of the policy rule. Johannes, Lochstoer, and Mou (2011) formally study the asset-pricing implication of Bayesian learning about a model's fundamental dynamics. Our focus is different. We note that the predictive content of lagged forward rates is a challenge for standard DTSMs where yield factors are Markovian and offer risk-based reconcilation with the data. Compensation for risk and informational frictions may both be consistent with non-Markovian dynamics in yields.

The class of Gaussian DTSMs includes the large family of VAR and VARMA

[3]Importantly, a hidden factor is not precluded. The generalization $\theta \neq 0$ also allows innovations with opposite loadings on the expectation term and the risk-premium term.

models, but empirical applications almost exclusively consider the VAR(1) case.[4] Monfort and Pegoraro (2007) consider cases with $0 < p < \infty$ lags and with regimes in the coefficients. We treat the case $p = \infty$ parsimoniously, circumventing the need for regimes (non-linearities).

Decomposing yields into a pure expectation component and the compensation for risk offered to bondholders is the *raison d'être* of modern DTSMs. The evidence of predictable bond return variations "stand[s] as challenges or 'stylized facts' to be explained by candidate models" (Fama, 1984).[5] Using the evidence in Campbell and Shiller (1991) as a benchmark, Dai and Singleton (2002) show that Gaussian DTSMs meet this challenge. But this literature ignores the information content of lagged forward rates. We extend the results in Dai and Singleton (2002), showing that conditional mean DTSMs capture the evidence of non-Markovian dynamics in bond risk premia.

In the context of macro-finance models, Ang and Piazzesi (2003) incorporate lags of macro variables (see also, e.g., Ang, Dong, and Piazzesi 2007; Jardet, Monfort, and Pegoraro 2008). More recently, Joslin et al. (2012) study macro-finance models where they project the economy-wide pricing kernel on the risk factors Z_t. Macro risks are unspanned in yields but help predict bond returns, consistent with the evidence in Ludvigson and Ng (2009). Economically, this is a restriction on the set of priced risk in the bond market. We see these approaches as complementary: both macro variables and past yields contain information about future returns. Feunou and Fontaine (2012) show the importance of including a moving-average term within a macro-finance term structure model that includes inflation. Joslin, Le, and Singleton (2013) consider lags of macro variables and the effect of spanning restrictions on estimated monetary policy rules.

The following section presents the evidence on bond returns predictability, emphasizing the empirical properties of $E[\mathcal{P}_{t+h}|\mathcal{I}_{\mathcal{P},t}]$. Section 2 develops the family of CM-DTSMs, and discusses in detail the connection with existing approaches. Section 3 reports all the results. All proofs are provided in the appendix.

[4] A common argument is that dynamics with longer lags can be recast within an extended state representation with only one lag. See, e.g., Equations (7)-(8) in Ang and Piazzesi (2003) or footnote 7 in Joslin, Singleton, and Zhu (2011). We document the empirical relevance of lagged yields.

[5] See also Shiller (1979); Startz (1982); Fama (1984); Fama and Bliss (1987); Campbell and Shiller (1991) for earlier evidence against the expectation hypothesis.

1 Evidence in Bond Risk Premia

This section builds on Cochrane and Piazzesi's 2005 (CP's) results and establishes a set of stylized facts to motivate a non-Markovian model for yields. Most importantly, we find that long lags of forward rates add to the predictability of bond returns. For each horizon, a single combination of current and past forward rates summarizes the predictability for bonds with different maturities. We show that (i) for each horizon, lagged coefficients decay exponentially, (ii) the return-forecasting factors are highly correlated across return horizons, and (iii) the predictability patterns are consistent with adding a parsimonious moving-average component to standard VAR(1) dynamics.

1.1 Cochrane-Piazzesi Regressions

CP document that a linear combination of forward rates forecasts annual excess returns from holding Treasury bonds. We repeat CP's analysis of returns predictability across holding periods between one month and one year. We consider the same predictors and the same sample period as in CP: annual forward rates with 1, 2, 3, 4 and 5 years to maturity from the Center for Research in Security Prices (CRSP) data set from 1963 until 2003. We find similar results in longer samples. We estimate the following regressions:

$$xr_{t,h}^{(n)} = b_{n,h}\gamma_h' f_t + u_{t,t+h}^{(n)}, \tag{1}$$

where $xr_{t,h}^{(n)}$ is the excess returns from holding a bond with maturity n over an h-month horizon, and where f_t stacks a constant with the forward rates. We compute excess returns for different horizons using the Gurkaynak et al. (2006) (GSW) data set.[6] Note that Equation (1) imposes the single-factor restriction: $b_{n,h}$ is a scalar and γ_h is a horizon-specific vector of coefficients. Panel (A) of Table 1 reports the R^2s for bond maturities up to 10 years. The predictability increases with the horizons – from around 3% at the 1-month horizon, to around 10%, 20%, 30% and 35% at the 3-, 6-, 9- and 12-month horizons, respectively. The last column matches results in CP. Unreported results show that the single-factor restriction is supported in the data: predictability regressions without a factor structure yield essentially the same results.

[6] In contrast, we can only compute annual excess returns from the CRSP Fama-Bliss data files.

1.2 Recursive Cochrane-Piazzesi Regressions

CP show that lags of the forward-based factor increase the predictability of annual returns. To illustrate, define $\bar{xr}_{t+12}$ as the cross-sectional average of annual excess returns across different bond maturities and extend CP's regression to include distributed lags of the forward rate factor. Allowing for eight lags,

$$\bar{xr}_{t+12} = \gamma' \left(a_0 f_t + \cdots + a_8 f_{t-8} \right), \tag{2}$$

the estimates $\hat{a}_j$ display a decaying pattern: 0.29, 0.25, 0.16, 0.14, 0.05, 0.07, 0.03 and 0.002.[7] The estimates are significant jointly but not individually, suggesting that a parsimonious model is required. Distributed-lag models with geometrically decaying coefficients have a long history in econometrics and arise naturally in partial-adjustment or adaptive-expectation models (see, e.g., Koyck 1954; Griliches 1967). We adapt a standard infinite distributed-lag model to the context of return predictability:

$$\begin{aligned} xr_{t,h}^{(n)} &= b_{n,h} \mathcal{R}_{t,h} + u_{t,h}^{(n)}, \\ \mathcal{R}_{t,h} &= (1-\alpha_h)\gamma_h' f_t + \alpha_h \mathcal{R}_{t-1,h}, \end{aligned} \tag{3}$$

where the scalar α_h is constant across maturities, and with $0 \leq \alpha_h < 1$, nesting Equation (1) if $\alpha_h = 0$. The weight on current forward rates is given by $1 - \alpha_h$, while α_h controls the weights on past forward rates. The lag representation of returns is given by

$$xr_{t,h}^{(n)} = b_{n,h}(1-\alpha_h)\gamma_h' \sum_{j=0}^{\infty} \alpha_h^j f_{t-j} + u_{t,h}^{(n)}. \tag{4}$$

Equation (3) parsimoniously captures the decaying patterns observed in the estimates from Equation (2). Beyond its simplicity, this reduced-form representation is also consistent with our term structure model (see Section 2.6).

Panel (B) of Table 1 reports the R^2s from estimates of Equation (3).[8] Relative to the case $\alpha_h = 0$, predictability increases for each horizon and maturity. The predictability almost doubles for the 1-month and 2-month horizons, from 3% to

[7] Compare with Table 5B in CP. We proceed via iterated OLS estimation over γ and a.

[8] Equation (3) can be estimated easily via non-linear least squares. We set the initial value $\mathcal{R}_{0,h}$ to the unconditional population mean. As in CP, we impose that $N^{-1}\sum_{n=1}^{N} b_{n,h} = 1$ to identify $b_{n,h}$ from γ_h and α_h where N is the number of maturity in the cross-section of excess returns.

6% and from 6% to around 11%, respectively, and increases substantially at longer horizons. Consistent with Duffee (2011), the relative importance of lagged forward rates increases as the returns horizon decreases.

Table 2 reports the coefficient estimates from Equation (3). For each horizon, the predictability gains follow from the scalar parameter α_h. Panel (A) shows that estimates of $\alpha_{n,h}$ exhibit a gradual decline across horizons, ranging between 0.8 and 0.6. Past forward rates are given a greater weight to forecast short-horizon returns. Panel (B) reports estimates for the loadings $b_{n,h}$. As expected, the estimates are centered around one and increase with the bond duration. Panel (C) reports estimates of γ_h. The tent shape is clear and strikingly similar even if estimation was carried out independently for each horizon.[9]

One natural question is whether we can collapse the return-forecasting factors $\gamma_h' f_t$ into an encompassing factor that does not depend on the horizon. Indeed, we find that the horizon-specific factors are highly correlated. In the case $\alpha = 0$, the first principal component extracted from the panel of return-forecasting factors (for different horizons) explains 93.7% of the total variations and the second component explains most of the remaining variations. For the more general case, Table 3 reports results from a principal-component analysis. The first principal component explains 97% of the total variation, with loadings that are spread out evenly across maturities.

1.3 Reconciling Yield Dynamics

What yield dynamics can connect the predictability evidence across these horizons? Following CP, define the yield vector $Y_t = [y_t^{(1)}\ y_t^{(2)}\ y_t^{(3)}\ y_t^{(4)}\ y_t^{(5)}]'$, where $y_t^{(n)}$ is the n-year zero-coupon yield, and consider a Markovian model for Y_t at the monthly frequency:

$$\Delta Y_t = K_{Y,0} + K_{Y,1} Y_{t-1} + \epsilon_{Y,t}, \tag{5}$$

where $\epsilon_{Y,t}$ is a mean-zero innovation. The Markovian assumption connects the distribution of future yields Y_{t+h} to current yields Y_t, exclusively. Importantly, lagged forward rates are excluded from the projection of future returns $xr_{t,t+h}^{(n)}$ on current forward f_t (e.g., $\alpha_h = 0$ in Equation (3)). But lagged forward rates play a significant role in the data. Should we add lagged yields Y_{t-j} to Equation (5)? For instance, CP add 12 lags to capture the predictability evidence from bond returns (for the purpose

[9] We also considered the unrestricted case where the forward rate coefficient $\gamma_{n,h}$ and the lag coefficient $\alpha_{n,h}$ vary with the horizon and the maturity of the bond. Estimates of the R^2s and the α remain essentially the same.

of constructing bootstrapped inference). But adding 11 lags to Equation (5) involves 275 parameters!

The data suggest a more direct solution. The pattern of decaying coefficients in Equations (2) and (3) may be due to a moving-average component in the yield dynamics. It is well known that inverting a moving-average term generates an autoregressive form with infinite lags, and with decaying coefficients (see, e.g, Hamilton 1994). Appendix A.1 shows that this pattern translates to the predictability coefficients. If a moving-average component is hidden in the residuals of Equation (5), then we have that

$$\epsilon_{Y,t} = u_t - \theta_Y u_{t-1}, \tag{6}$$

where θ_Y is a matrix and u_{t-1} is white noise. A direct implication is that successive residuals $\epsilon_{Y,t}$ and $\epsilon_{Y,t-1}$ are correlated. Hence, we estimate the VAR(1) in Equation (5) and we assess the correlation structure in the panel of residuals.

First, we focus on the time-series dimension. Panel (A) of Table 4 reports the results from regressions of each element of $\epsilon_{Y,t}$ on its own lag. We find that the first lag is significant in the majority of cases, but that longer lags are always insignificant.[10] Next, we focus on the first lag and examine the cross-sectional correlations. Panel (B) shows estimates of the cross-correlation between each element of $\epsilon_{Y,t}$ and each element of its lag $\epsilon_{Y,t-1}$. Again, the individual correlation coefficients are significant at the 5% level in the majority of cases, and significant at the 10% level in almost all cases. In addition, the matrix of estimates in Panel (B) reveals a striking factor structure: each column is nearly a parallel shift of the previous column. In other words, each component of $\epsilon_{Y,t}$ is correlated with the same linear combination of $\epsilon_{Y,t-1}$.

This factor structure implies that a regression of one element $\epsilon_{Y,t}^{(n)}$ on the vector $\epsilon_{Y,t}$ is inappropriate due to the collinearity of the coefficient matrix. We adopt the following two-step estimation procedure. First, regress the cross-section average $\bar{\epsilon}_{Y,t} \equiv N^{-1} \sum \epsilon_{Y,t}^{(n)}$, with $N = 5$, on $\epsilon_{Y,t-1}$:

$$\bar{\epsilon}_{Y,t} = \gamma_\epsilon^\top \epsilon_{Y,t-1} + \bar{w}_t,$$

and then regress each element $\epsilon_{Y,t}^{(n)}$ on $\hat{\gamma}_\epsilon^\top \epsilon_{Y,t-1}$. These two steps provide an estimate of the single-factor regression:

$$\epsilon_{Y,t} = \alpha \gamma_\epsilon^\top \epsilon_{Y,t-1} + w_t, \tag{7}$$

[10] We cannot recover a consistent estimate of θ based on the residuals from the VAR1 $\hat{\epsilon}_{Y,t+1} = \Delta Y_{t+1} - \hat{K}_{Y,0} - \hat{K}_{Y,1} Y_t$, since the estimate $\hat{K}_{Y,1}$ is based on a misspecified model.

with the identification restriction that $\sum_n \alpha_n = N$.[11] Panels (C) and (D) of Table 4 report the results from the first step and the second step, respectively. The estimate of γ_{ϵ_t} shows the familiar tent shape. Individual t-statistics are low due to the collinearity, but we can reject the null hypothesis that the vector of coefficients is jointly equal to zero: the F-statistic for restriction $\gamma_{\epsilon_t} = 0$ has a p-value of 0.02. Finally, the coefficient estimates α_n on the lagged factor $\gamma_\epsilon^\top \epsilon_{Y,t-1}$ are all statistically significant.

The predictability evidence and the time-series evidence both support the presence of a moving average. To conclude, and before turning to the formal modelling section, we sketch how we can connect the two pieces of evidence. The simplest way to introduce a moving-average component assumes that the conditional mean $\mathcal{E}_{Y,t} \equiv E[Y_{t+1}|\mathcal{I}_{Y,t}]$ is Markovian:

$$\begin{aligned} Y_t &= \mathcal{E}_{Y,t-1} + \Sigma_Y \epsilon_{Y,t} \\ \Delta\mathcal{E}_{Y,t} &= K_{Y,0} + K_{Y,1}\mathcal{E}_{Y,t-1} + \Sigma_{\mathcal{E}_Y}\epsilon_{Y,t}, \end{aligned} \tag{8}$$

where ϵ_t is white noise. The benchmark is the VAR(1) and we also estimate a VAR(12), as in CP. For comparability, we estimate each model via least-squares. Using the estimates, we then derive the model-implied population R^2s for the regression in Equation (1). Table 5 reports the results. Consistent with CP's results, the predictability of annual returns implied by the VAR(1) model is typically half that obtained from direct regressions. On the other hand, the VAR(12) implies population R^2s that are close to the regression results. Strikingly, adding a small moving-average component also matches the evidence. Unsurprisingly, this suggests that the VAR(12) is overparameterized.

2 Markovian Yields and Measurement Errors

The previous section revisited and added to the evidence showing that lagged forward rates predict bond excess returns. CP suggest that lagged forward rates are informative because: "time-t yields (or prices, or forwards) truly are sufficient state variables, but there are small measurement errors that are poorly correlated over time." Then, it follows from the standard Kalman "that the best guess of the true [state] is a geometrically weighted moving average." (See Cochrane and Piazzesi 2005, p. 154.) This is in the same spirit as Duffee's 2011 advocacy of a significant change in the construction of multifactor term structure models. Specifically, multifactor models

[11] CP use a similar two-step procedure in predictability regressions with a single-factor structure.

must allow one factor to determine investors' expectations of future yields, but with no effect on current yields. Arguably, Duffee's hidden factor has two drawbacks: (i) this factor must have opposite effects on expected future interest rates and on bond risk premia *at all times and for every yield maturity*, and (ii) the hidden factor is only detectable in a term structure model with five sources of risk.

This section shows that none of these features is needed to generate a role for lagged forward rates. We construct a parsimonious extension of the standard Markovian DTSMs where the current cross-section of yields is not a sufficient state variable to forecast future yields. We differ from Duffee (2011) and Joslin, Priebsch, and Singleton (2012) in that we do not rely on any price of risk restriction to generate the hidden factor. However, we follow Duffee (2011) in allowing for measurement errors in all yields.

2.1 Risk-Neutral Dynamics

We postulate a small number of $\mathcal{N}$ stationary Gaussian risk factors Z_t driving the cross-section of bond yields:

Assumption 1 *The cross-section of n-period yields, $y_t^{(n)}$, $n \geq 1$, can be expressed as a linear function of Z_t,*

$$y_t^{(n)} = A_n + B_n' Z_t, \tag{9}$$

where $Z_t \in \mathbb{R}^{\mathcal{N}}$ has a stationary Gaussian distribution.

Assumption 1 is consistent with the observation that a small number of principal components suffice to explain most of the variations in the term structure of yields. Note that the coefficients A_n and B_n are free at this stage. Cross-equation restrictions arise if we assume that bond prices offer no arbitrage opportunity. Proposition 1 clarifies an important implication of the absence of arbitrage.

Proposition 1 *If bond prices offer no arbitrage opportunity, Assumption 1 implies that Z_t has Markovian dynamics under $\mathbb{Q}$ and we can write*

$$\Delta Z_t = K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}} Z_{t-1} + \Sigma \epsilon_{Z,t}^{\mathbb{Q}}, \tag{10}$$

where $\epsilon_{z,t}^{\mathbb{Q}}$ is a standard Gaussian innovation. Then, A_n and B_n' follow (standard) recursions, which are given in the appendix.

Proposition 1 is useful in clarifying that once we fix $\mathcal{N}$, the number of linear risk factors required to explain the cross-section of yields, then these factors must have linear Markovian dynamics under risk-neutral dynamics. The result is not trivial but depends crucially on the absence of arbitrage opportunities among bond prices. Nonetheless, this result should not come as a surprise. Almost all existing models consider the natural case where Z_t has Markovian dynamics under the risk-neutral measure.

2.2 Historical Dynamics

If Z_t is Markovian under $\mathbb{Q}$, how is the no-arbitrage restriction consistent with the stylized fact that *past* yields predict excess returns? To answer this question, consider the excess returns from holding an n-period bond between t and $t+h$. Under the conditions of Proposition 1, this is given by

$$xr_{t,h}^{(n)} = -\frac{h}{2}\mathcal{B}'_{n-1}\Sigma\Sigma'\mathcal{B}_{n-1} + \mathcal{B}'_{n-1}\sum_{j=1}^{h}\left(\Delta Z_{t+j} - E^{\mathbb{Q}}\left[\Delta Z_{t+j}|\mathcal{I}_{Z,t+j-1}\right]\right), \tag{11}$$

where $\mathcal{B}_{n-1} = -nB_n$. Taking conditional expectations under the $\mathbb{P}$-measure, $E^{\mathbb{P}}[xr_{t,h}^{(n)}|\mathcal{I}_{Z,t}]$, each summand in the second term becomes

$$E^{\mathbb{P}}\left[-K_0^{\mathbb{Q}} + (I_{\mathcal{N}} - K_1^{\mathbb{Q}})\Delta Z_{t+j}|\mathcal{I}_{Z,t+j-1}\right], \tag{12}$$

where we use the fact that the time-t conditional expectation operator under $\mathbb{Q}$ is a function of Z_t only. Therefore, Proposition 1 implies that the predictability of bond excess returns in Equation (11) depends on lagged yields if and only if the yield factors Z_t are non-Markovian under the $\mathbb{P}$-measure. This can arise from two distinct mechanisms. Conditional on the econometrician's information set, this expectation will be a function of the past if we observe Z_t with errors. Alternatively, the $\mathbb{P}$-dynamics of Z_t could be non-Markovian from the standpoint of investors who observe the risk factors directly. We allow for both effects. Proposition 2 introduces CM dynamics under $\mathbb{P}$.

Assumption 2 *The risk factors $Z_t \in \mathbb{R}^{\mathcal{N}}$ have the following generic conditional mean dynamics:*

$$\begin{aligned} Z_t &= \mathcal{E}_{Z,t-1}^{\mathbb{P}} + \Sigma\epsilon_t^{\mathbb{P}} \\ \Delta\mathcal{E}_{Z,t}^{\mathbb{P}} &= K_0^{\mathbb{P}} + K_1^{\mathbb{P}}\mathcal{E}_{Z,t-1}^{\mathbb{P}} + \Sigma_{\mathcal{E}_Z}\epsilon_t^{\mathbb{P}}, \end{aligned} \tag{13}$$

where $\mathcal{E}^{\mathbb{P}}_{Z,t} \equiv E^{\mathbb{P}}[Z_{t+1}|\mathcal{I}_{Z,t}]$ *and* $\epsilon^{\mathbb{P}}_t$ *is a standard Gaussian innovation. Define*

$$\theta \equiv I_{\mathcal{N}} + K^{\mathbb{P}}_1 - \Sigma_{\mathcal{E}_Z}\Sigma^{-1}. \tag{14}$$

Then, Z_t *has Markovian dynamics if and only if* $\theta = 0$.

Spanning vs. forecasting

Relaxing the Markovian assumption for Z_t implies that the number of state variables increases from $\mathcal{N}$ to $2\mathcal{N}$, but with no increase in the sources of risk. The states Z_t and $\mathcal{E}^{\mathbb{P}}_{Z,t}$ are driven by the *same* $\mathcal{N}$ shocks $\epsilon^{\mathbb{P}}_t$. The Z_t innovations are given by $\Sigma\epsilon^{\mathbb{P}}_t$, while the $\mathcal{E}^{\mathbb{P}}_{Z,t}$ innovations are given by $\Sigma_{\mathcal{E}_Z}\epsilon^{\mathbb{P}}_t$. The joint covariance matrix for Z_t and $\mathcal{E}^{\mathbb{P}}_{Z,t}$ has rank $\mathcal{N}$ only (conditionally or unconditionally as estimated via principal-component analysis, say). This is uncontroversial, since the same properties hold in a VAR where $\mathcal{E}^{\mathbb{P}}_{Z,t}$ is a deterministic function of Z_t.

Although they are driven by the same shocks, the states Z_t and $\mathcal{E}_{Z,t}$ play distinct roles. The yield factors Z_t summarize the cross-section of yields: there are $\mathcal{N}$ principal components in the cross-section of yields. On the other hand, the conditional means $\mathcal{E}_{Z,t}$ summarize the future evolution of yields and it is a function of the entire history of Z_t. To see this, iterate Equation (13) backward. This separation of the *spanning* and the *forecasting* roles is at the heart of our modelling strategy.

Markovian Expectations

The conditional mean $\mathcal{E}^{\mathbb{P}}_{Z,t}$ is Markovian under $\mathbb{P}$, but the risk factors Z_t are not. The conditional expectations $\mathcal{E}^{\mathbb{P}}_{Z,t}$ depend on lagged expectations, $\mathcal{E}^{\mathbb{P}}_{Z,t-1}$, via the auto-regressive matrix $(I_{\mathcal{N}} + K^{\mathbb{P}}_1)$, and on the current shock $Z_t - \mathcal{E}^{\mathbb{P}}_{Z,t-1}$, via the matrix $\Sigma_{\mathcal{E}_Z}\Sigma^{-1}$. The restriction $\theta = 0$ states that Z_t is Markovian when these weights are equal. Consistent with the evidence in Section 1, Assumption 2 is equivalent to an unrestricted VARMA(1,1) for Z_t where θ is the moving-average parameter.[12] Therefore, the dependence between $\mathcal{E}^{\mathbb{P}}_{Z,t}$ and the past is characterized parsimoniously by θ.[13] Note that the conditional mean $\mathcal{E}^{\mathbb{P}}_{Z,t}$ can be filtered easily given the time-series of $Z^{\mathbb{P}}_t$ using the recursion in Equation (13) and starting from some initial value $\mathcal{E}^{\mathbb{P}}_{Z,0}$. Nonetheless, Z_t is latent in our context and must be filtered from its effect on yields.

[12] This is analogous to a GARCH(1,1) model where the conditional variance can be filtered directly from observed returns. The general CM representation was introduced in Fiorentini and Sentana (1998) in the context of time-series models.

[13] Alternatively, one can directly introduce lagged values of Z_t in a VAR(p) representation. We argue that this approach is not parsimonious and that the conditional mean representation is more intuitive in the context of term structure models (where conditional expectations play a key role).

General Equilibrium

Non-Markovian effects arise naturally within DSGE models. Expectations and current values are driven by the same set of shocks; see for instance the discussion in Fernández-Villaverde et al. (2007). In addition, Ravenna (2007) discusses the invertibility problem associated with finding a finite-order VAR representation in DSGE models. Non-Markovian effects also arise in a limited-information context even if there exist some Markovian state variables. Consider an equilibrium where Z_t is part of a broader system involving other financial or macro variables, say $\Xi_t \equiv [Z_t \; Z_{2,t}]$, and where Ξ_t is Markovian and stationary. Then, the subset Z_t, grouping only those variables that explain the cross-section of yields, will in general have non-Markovian dynamics under $\mathbb{P}$, unless specific exogeneity conditions are imposed between Z_t and $Z_{2,t}$.

2.3 Conditional Mean Representation of the Kalman Filter Dynamics

To complete the specification of our model, we specify the statistical properties of observed yields. Let $\{m_1, m_2, \ldots, m_{\mathcal{J}}\}$ be the set of maturities (in years) of the bonds used in estimation, $Y_t \equiv \{y_t^{(m_1)}, \ldots, y_t^{(m_{\mathcal{J}})}\} \in \mathbb{R}^{\mathcal{J}}$ be the set of model-implied yields (see Equation (9)) and Y_t^o the corresponding set of observed yields. As above, define $\mathcal{I}_{\mathcal{P},t} = \{\mathcal{P}_t^0, \mathcal{P}_{t-1}^o, \ldots\}$, the history of $\mathcal{P}_t^o$. We consider the case where measurement errors are pervasive.

Assumption 3 *Any yield portfolio $\mathcal{P}_t^o \equiv WY_t$, with W a $\mathcal{J} \times \mathcal{J}$ matrix of portfolio weights, equals its DTSM-implied values $\mathcal{P}_t \equiv WY_t^o$ plus a mean zero, independent and normally distributed error, $w_t = \mathcal{P}_t^o - \mathcal{P}_t$.*

Assumption 3 leads to a well-known Kalman filtering problem, which is often summarized via the following state-space representation:

$$\mathcal{P}_t^o = H_0 + H' Z_t + w_t \tag{15}$$

$$Z_t = \mathcal{E}_{Z,t-1}^{\mathbb{P}} + \Sigma \epsilon_t^{\mathbb{P}}. \tag{16}$$

Equation (15) stacks the measurement equations for yields and Equation (16) corresponds to the state dynamics. This representation is based on the latent factors Z_t: all time-series and cross-section implications from the model are not direcly available. It is only after a recursive application of the Kalman filter to the data that model forecasts and model yields can be obtained using the filtered estimates of Z_t and $\mathcal{E}^{\mathbb{P}}_{Z,t-1}$. This also delivers the likelihood of the data $\{\mathcal{P}^o_t\}$ for the purpose of parameter estimation.

Hence, all the relevant model implications of the model must be derived in terms of the observables $\mathcal{P}^o_t$. Theorem 1 establishes a central result: the CM-DTSM given by Equations (15)-(16) can be represented directly in terms of the observed $\mathcal{P}^o_t$ and its conditional mean $\mathcal{E}_{\mathcal{P},t}$. The resulting model belongs to the family of CM-DTSMs: the portfolio dynamics is analogous to that in Assumption 2. This holds whether Z_t has CM dynamics or $\mathcal{P}^o_t$ are measured with errors.

Theorem 1 *Let $\mathcal{P}_t = H_0 + H'Z_t$ for some $\mathcal{J} \times 1$ vector H_0 and $\mathcal{N} \times \mathcal{J}$ full-row rank matrix H, where $\mathcal{J} > \mathcal{N}$. Define $R \equiv var(\mathcal{P}_t - \mathcal{P}^o_t)$, a full-rank diagonal matrix, and $\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t} \equiv E^{\mathbb{P}}[\mathcal{P}^o_{t+1}|\mathcal{I}_{\mathcal{P},t}]$. Then, $\mathcal{P}^o_t$ has the following conditional mean dynamics:*

$$\mathcal{P}^o_t = \mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1} + \Sigma_{\mathcal{P}}\epsilon^{\mathbb{P}}_{\mathcal{P},t} \tag{17}$$

$$\Delta\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t} = K^{\mathbb{P}}_{0\mathcal{P}} + K^{\mathbb{P}}_{1\mathcal{P}}\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1} + \Sigma_{\mathcal{E}_{\mathcal{P}}}\epsilon^{\mathbb{P}}_{\mathcal{P},t}. \tag{18}$$

Our canonical form is parameterized by $\Theta^{\mathcal{P}} = \{K^{\mathbb{P}}_{0\mathcal{P}}, H, K^{\mathbb{P}}_1, R, \theta\}$ with positive elements on the first line of H and with the eigenvalues of $K^{\mathbb{P}}_{1\mathcal{P}}$ ordered from high to low. Define $H^{\perp}H' = I_{\mathcal{N}}$. The other parameters in Equations (17)-(18) *are given by*

$$\begin{aligned} K^{\mathbb{P}}_{1\mathcal{P}} &= H'(K^{\mathbb{P}}_1 + I_{\mathcal{N}})H^{\perp} - I_{\mathcal{J}} \\ K^{\mathbb{P}}_{0\mathcal{P}} &= (I_{\mathcal{J}} - K^{\mathbb{P}}_{1\mathcal{P}})H_0 + H'K^{\mathbb{P}}_0 \\ \Sigma_{\mathcal{E}_{\mathcal{P}}} &= \left((K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}})P - H'\theta H\right)(\Sigma'_{\mathcal{P}})^{-1} \\ \Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}} &= P + R, \end{aligned} \tag{19}$$

where P has rank $\mathcal{N}$ and is given as the solution of a generalized algebraic Ricatti equation in terms of the canonical parameters (see Appendix A.3). Define

$$\theta_{\mathcal{P}} \equiv H'(K^{\mathbb{P}}_1 + I_{\mathcal{N}})H^{\perp}R(P+R)^{-1} + H'\theta H(P+R)^{-1}. \tag{20}$$

Then, $\mathcal{P}^o_t$ is non-Markovian if and only if $\theta_{\mathcal{P}} \neq 0$.

We refer to the CM-DTSM in Theorem 1 as the *CM* canonical form parameterized by $\Theta^{\mathcal{P}}$.[14] One identification assumption is that $\Sigma = I_{\mathcal{N}}$. Theorem 1 summarizes all

[14] It is tempting to stack Z_t and $\mathcal{E}_{Z,t}$ within an extended VAR representation and apply the results in Joslin, Singleton, and Zhu (2011), but these are not applicable: their canonical form requires an unrestricted VAR specification.

the dynamic implications of the model. This is really all we can know about the yield dynamics given the observable data $\mathcal{I}_{\mathcal{P},t}$. It is similar in spirit to Theorem 1 in Joslin, Singleton, and Zhu (2011) deriving a canonical model in terms of yield portfolios measured without error. Our proof connecting the state-space representation in Equations (15)-(16) to the direct representation in Equations (17)-(18) uses the steady-state solution of the Kalman filter. This is innocuous in our case – the distribution is Gaussian with constant variance – since the filter converges very fast to its steady state irrespective of the initial gain matrix.

2.4 Two Mechanisms for Unspanned Information

Equation 20 identifies two mechanisms that generate non-Markovian effects in the dynamics of the portfolios $\mathcal{P}_t^o$. This is one clear benefit from the representation in Theorem 1. Rearranging Equation (20), $\theta_{\mathcal{P}}$ is given by

$$\theta_{\mathcal{P}} = \left(H'(K_1^{\mathbb{P}} + I_{\mathcal{N}})H^{\perp}\right) R(\Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}')^{-1} + \left(H'\theta H^{\perp}\right) Q_{\mathcal{P}}(\Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}')^{-1}, \tag{21}$$

where $H'(K_1^{\mathbb{P}}+I_{\mathcal{N}})H^{\perp}$ is the persistence of $\mathcal{P}_t$, $(H'\theta H^{\perp})$ the moving-average coefficient and $Q_{\mathcal{P}} = H'\Sigma\Sigma'H$ the covariance matrix of its innovations. The portfolios $\mathcal{P}_t^o$ are not Markovian, $\theta_{\mathcal{P}} \neq 0$, whether the latent risk factor Z_t is non-Markovian, $\theta \neq 0$, or the portfolios are measured with errors, $R = 0$, or both.

The first term in Equation (21) is zero if $R = 0$. This term is the source of the hidden factor in Duffee (2011): the conditional dynamics of $\mathcal{P}_t^o$ depends on the history of yields via the Kalman filter. This is why the estimate of $\theta_{\mathcal{P}}$ is closely related to the estimate of the yield persistence $H'(K_1^{\mathbb{P}} + I_{\mathcal{N}})H^{\perp}$. Empirically, the yield dynamics implied by the Kalman filter should be close to Markovian if the measurement errors are small relative to the innovations in yield portfolios; i.e., if R is "small" relative to $\Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}'$.[15] This mechanism is further restricted, since R is a diagonal covariance matrix ($\mathcal{N}$ parameters), to preserve the interpretation of the measurement errors. The more restricted case $R = \sigma^2 I$ is often used in practice, as in Duffee (2011).

The second term in Equation (21) provides a more flexible channel to generate non-Markovian dynamics. Since measurement errors are small, the innovations in $\mathcal{P}_t$ represent a large share of the innovations in $\mathcal{P}_t^o$ and $Q_{\mathcal{P}}$ is close to $\Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}'$. Therefore, the moving-average term in the risk factor Z_t is communicated directly to the yield.[16]

[15] The dynamics will also be close to Markovian if the persistence of the yield portfolios $(K_{1\mathcal{P}}^{\mathbb{P}}+I_{\mathcal{J}})$ is small, but this case is irrelevant.

[16] Equation (21) suggests that θ and R interact in the determination of $\theta_{\mathcal{P}}$: θ appears in the first

This mechanism differs from that for the unspanned yield factors in Duffee (2011) or for the unspanned macro variables in Joslin, Priebsch, and Singleton (2012). As discussed above, $\mathcal{E}_t$ and Z_t have different roles if the latent factor Z_t is non-Markovian. It suffices that $\theta \neq 0$; then, Z_t spans the cross-section of yields but current yields do not span the conditional dynamics of Z_t. There is no need to restrict the pricing kernel.

2.5 Risk-Neutral Parameters

The canonical form in Theorem 1 does not include parameters of the $\mathbb{Q}$-dynamics for Z_t. Instead, we allow for free parameters in H to emphasize that the result does not rely on the no-arbitrage assumption. Nonetheless, this structure imposes cross-equation restrictions from the generic factor model $\mathcal{P}_t = H_0 + H' Z_t$. If, in addition, bond prices are free of arbitrage opportunity, the vector H_0 and the matrix H embody additional restrictions that are derived from the $\mathbb{Q}$-dynamics of Z_t:

$$H_0 = \begin{bmatrix} A_{n_1} \\ \vdots \\ A_{n_\mathcal{J}} \end{bmatrix} \qquad H = \begin{bmatrix} B'_{n_1} \\ \vdots \\ B'_{n_\mathcal{J}} \end{bmatrix}, \tag{22}$$

where A_n and B'_n follow standard recursions, which are given in terms of $K_0^{\mathbb{Q}}$, $K_1^{\mathbb{Q}}$, Σ, as well as $\delta_0 \equiv A_1$ and $\delta_1 \equiv B_1$. Using a standard identification assumption for the risk-neutral parameters generates the following alternative canonical form for the dynamics of $\mathcal{P}_t^o$:

$$\widetilde{\Theta}^{\mathcal{P}} = \{\delta_0, \delta_1, K_1^{\mathbb{Q}}, K_1^{\mathbb{P}}, R, \theta\}, \tag{23}$$

where $\delta > 0$ (element by element). The $\mathcal{N} \times (\mathcal{N}+1)$ risk-neutral parameters δ_1 and $K_1^{\mathbb{Q}}$ replace the $\mathcal{N} \times \mathcal{J}$ matrix H (recall that $\mathcal{J} > \mathcal{N}$). The connection between δ_1, $K_1^{\mathbb{Q}}$ and H is given directly from Equation (22) and Proposition 1. We set $K_1^{\mathbb{Q}} = 0$ for identfication. The matrix $K_{0\mathcal{P}}^{\mathbb{P}}$ is not included in $\widetilde{\Theta}^{\mathcal{P}}$, but it is given by

$$K_{0\mathcal{P}}^{\mathbb{P}} = (I_\mathcal{J} - K_{1\mathcal{P}}^{\mathbb{P}}) H_0 + H' K_0^{\mathbb{P}},$$

in terms of other parameters (see Theorem 1) with H_0 given in Equation (22).

The pricing equation for all the observed portfolios $\mathcal{P}_t^o$ is embedded within Equations (17)-(18). Conditioning on $\mathcal{I}_{\mathcal{P},t}$, the pricing equation $y_t^{(n)} = A_n + B'_n Z_t$ for any

term, and R appears in the second term. However, this effect is of second order and can be safely ignored in practice.

other maturity n becomes

$$y_t^{(n)} = A_{\mathcal{P},n} + B'_{\mathcal{P},n}\mathcal{P}_t^o + C'_{\mathcal{E}_\mathcal{P},n}\mathcal{E}_{\mathcal{P},t}, \tag{24}$$

with coefficients

$$\begin{aligned} A_{\mathcal{P},n} &= A_n - H^\perp H_0 \\ B_{\mathcal{P},n} &= B'_n H^\perp P(P+R)^{-1} \\ C_{\mathcal{P},n} &= B'_n H^\perp (I_\mathcal{J} - P(P+R)^{-1}). \end{aligned} \tag{25}$$

2.6 Bond Risk Premium

This section shows that the same mechanisms generating unspanned risk factors also generate lagged coefficients in the bond risk premium. For that purpose, Proposition 2 first derives the pricing kernel that is consistent with Equations (10) and (13), respectively.

Proposition 2 *The unique pricing kernel M_{t+1} consistent with the $\mathbb{Q}$-dynamics and the $\mathbb{P}$-dynamics in Equations* (10) *and* (13)*, respectively, is given by*

$$M_{t+1} \equiv \exp\left(-\frac{\Lambda'_t \Sigma\Sigma' \Lambda_t}{2} - \Lambda'_t \Sigma \epsilon^\mathbb{P}_{t+1}\right),$$

with prices of risk

$$\Lambda_t \equiv (\Sigma\Sigma')^{-1}\left(\Lambda_0 + \Lambda_Z Z_t + \Lambda_\mathcal{E}\mathcal{E}^\mathbb{P}_{Z,t-1}\right). \tag{26}$$

The mapping between parameters of Equations (10) *and* (13) *is given by*

$$\begin{aligned} K_0^\mathbb{P} &= \Lambda_0 + K_0^Q \\ K_1^\mathbb{P} &= \Lambda_Z + \Lambda_\mathcal{E} + K_1^\mathbb{Q} \\ \Sigma_{\mathcal{E}_Z} &= \left(\Lambda_Z + K_1^\mathbb{Q} + I_\mathcal{N}\right)\Sigma. \end{aligned} \tag{27}$$

Therefore, the CM dynamics for Z_t in Assumption 2 is a generalization of the prices of risk. Indeed, we have that $\theta = 0$ is equivalent to $\Lambda_\mathcal{E} = 0$, yielding the standard pricing kernel. Otherwise, the prices of risk are functions of both Z_t and $\mathcal{E}_{Z,t-1}$. More generally, $\Lambda_\mathcal{E} \neq 0$ is consistent with a habit specification (Campbell and Cochrane, 1999). It is also consistent with a moving-average component in the state dynamics of a long-run risk economy (Bansal and Yaron, 2000). Building on this result, Proposition 3 confirms that long lags of Z_t enter the bond risk premium in this non-Markovian market.

Proposition 3 *For CM-DTSMs, the risk premium from holding an n-period bond between t and t + h,*

$$brp_{t,h}^{(n)} \equiv E[xr_{t,h}^{(n)}|\mathcal{I}_{Z,t}], \tag{28}$$

is given by

$$brp_{t,h}^{(n)} = \delta_{h,0} + \delta_{h,Z}^{(n)\prime} Z_t + \delta_{h,\mathcal{E}}^{(n)\prime} \mathcal{E}_{Z,t-1}^{\mathbb{P}}, \tag{29}$$

where $xr_{t,h}^{(n)}$ is defined in Equation (11). We have that $\delta_{h,\mathcal{E}}^{(n)} = 0$ if and only if Z_t is Markovian under $\mathbb{P}$ ($\theta = 0$).

The coefficients $\delta_{h,Z}^{(n)}$ and $\delta_{h,\mathcal{E}}^{(n)}$ in Equation (29) are given by

$$\begin{aligned}
\delta_{h,Z}^{(n)\prime} &= B_{n-1}' \times \left[\Lambda_Z - (\Lambda_Z + \Lambda_{\mathcal{E}}) \left(K_1^{\mathbb{P}}\right)^{-1} \left(I_{\mathcal{N}} - \left(K_1^{\mathbb{P}} + I_{\mathcal{N}}\right)^{h-1}\right) \Sigma_{\mathcal{E}} \Sigma^{-1}\right] \\
\delta_{h,\mathcal{E}}^{(n)\prime} &= B_{n-1}' \times \left[I_N + (\Lambda_Z + \Lambda_{\mathcal{E}}) \left(K_1^{\mathbb{P}}\right)^{-1} \left(I_N - \left(K_1^{\mathbb{P}} + I_N\right)^{h-1}\right)\right] \Lambda_{\mathcal{E}}.
\end{aligned} \tag{30}$$

Again, we cannot use Equation (29) directly in practice, since Z_t is latent (and therefore $\mathcal{E}_{Z,t}^{\mathbb{P}}$) and must be filtered. Conditioning on $\mathcal{I}_{\mathcal{P},t}$, the expected bond risk premium is given by

$$\begin{aligned}
\widehat{brp}_{t,h}^{(n)} &= E[brp_{t,h}^{(n)}|\mathcal{I}_{\mathcal{P},t}] \\
&= \delta_{h,0\mathcal{P}} + \delta_{h,Z}^{(n)\prime} H^{\perp} P(P+R)^{-1} \mathcal{P}_t^o \\
&\quad + \left[\delta_{h,Z}^{(n)\prime} H^{\perp} \left(I - P(P+R)^{-1}\right) + \delta_{h,\mathcal{E}}^{(n)\prime} H^{\perp}\right] \mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}}.
\end{aligned} \tag{31}$$

The risk premium depends on the current forward rates, via the yield portfolios $\mathcal{P}_t^o$, but also depends on past forward rates, via $\mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}}$. Again, there are two channels driving the results, as shown by the two terms in brackets. Past forward rates will help predict the future bond premium whether $\Lambda_{\mathcal{E}} \neq 0$ (in which case $\delta_{h,\mathcal{E}}^{(n)} \neq 0$) or $R \neq 0$. The role of the history of yields in the $\mathbb{P}$-dynamics and its role in the bond risk premium are intertwined. Therefore, the CM model is consistent with the predictability evidence: Equation (31) is similar to the reduced-form specification in Equation (4).

2.7 Unspanned Macro Risks

Joslin, Priebsch, and Singleton (2012) generate unspanned information via a very different mechanism. To see the difference, start with a simple Markovian model for the yield factor Z_t and expand the system with one latent factor $Z_{2,t}$. This additional variable affects the future of Z_t and it comes with its own shock $\epsilon_{2,t}$. Then, the spanning restriction in Joslin, Priebsch, and Singleton (2012) corresponds to a projection of the economy-wide pricing kernel, which is a function of ϵ_t and $\epsilon_{2,t}$ on "the priced risks in the bond market and on the state of the economy." Only then is the extra factor $Z_{2,t}$ not spanned by the current yield curve.

This restriction connects the parameters governing the prices of the risks ϵ_t and $\epsilon_{2,t}$ with the parameters governing the dynamics of Z_t and $Z_{2,t}$ under the historical measure $\mathbb{P}$ (see Appendix B in Joslin, Priebsch, and Singleton 2012). In contrast, we relax the Markovian assumption. This introduces a role for lagged forward rates directly, consistent with the bond premium evidence in Section 1, but does not require additional sources of risk. Hence, we do not need to project the pricing kernel M_t on a subset of the sources of risk and there is no need to connect the parameters driving the prices of ϵ_t with those driving the dynamics of Z_t.[17]

3 Results

This section studies the roots of the non-Markovian bond risk premium. We explore two alternative channels in great detail, asking the following questions. First, are the risk factors truly Markovian? In this case, small measurement errors should be sufficient to generate the predictability seen in the data. Second, if not, can a small number of non-Markovian risk factors with conditional mean dynamics match the evidence? Pre-empting the results, we conclude that measurement errors on their own are not sufficient to match the evidence, but that a CM model with three factors matches the evidence.

3.1 Nomenclature

We use the following model nomenclature. Each model is designated by a label of the form $CM_{\mathcal{N}}^{n}$-$KF\mathcal{M}$. The subscript $\mathcal{N}$ indicates the number of latent factors, and the superscript $n \leq \mathcal{N}$ indicates the rank of θ: the case $n = 0$ corresponds to the case where $\theta = 0$ and the risk factors are Markovian. Finally, $\mathcal{M}$ designates different structures for the covariance matrix of measurement errors R. We consider three cases. $KF0$ supposes that $\mathcal{N}$ linear combinations of the $\mathcal{J}$ yields are observed

[17]This should make clear why using a stochastic mean model for Z_t, where the conditional mean has its own shocks, cannot resolve the tension between the time-series and the cross-sectional properties.

without errors. In this case, we invert the yield equations to reveal the risk factors, following standard practice. The other two cases suppose that all yields are measured with errors. $KF1$ designates models where $R = \sigma^2 I_{\mathcal{N}}$ and $KF2$ designates models where R is a free diagonal matrix.

3.2 Data

We use end-of-month zero-coupon yields from CRSP between December 1963 and December 2007 and focus on bonds with annual maturities of 1 to 5 years. This choice of sample period and maturities eases comparison with the results in Cochrane and Piazzesi (2005) and Duffee (2011). For instance, the evidence of a non-Markovian bond premium in Cochrane and Piazzesi (2005) is based on a monthly Markovian model for the first five annual forward rates. Similarly, Duffee (2011) uses these yields to argue for the presence of a hidden factor. As in Cochrane and Piazzesi (2005) , we set $\mathcal{P}_t = [f_t^{(12)}\ f_t^{(24)}\ f_t^{(36)}\ f_t^{(48)}\ f_t^{(60)}]'$ where $f_t^{(n)}$ is the 1-year forward rate $n/12 - 1$ years ahead ($f_t^{(1)}$ is the short rate).

3.3 Likelihood

The extant literature uses the Kalman filter for estimation whenever all the portfolios $\mathcal{P}_t^o$ are measured with errors. Theorem 1 shows the equivalence between the steady-state Kalman filter and the $CM_{\mathcal{N}}^k$-$KF1$ and $CM_{\mathcal{N}}^k$-$KF2$ models. Then the joint likelihood of yields observed at time t is given by

$$f(\mathcal{P}_t^o|\mathcal{I}_{t-1};\Theta^{\mathcal{P}}) = -\frac{1}{2}\left(\log \Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}' + (\mathcal{P}_t^o - \mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}})'(\Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}')^{-1}(\mathcal{P}_t^o - \mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}})\right), \quad (32)$$

where $\mathcal{I}_{t-1}$ includes the history of observed portfolios $\{\mathcal{P}_{t-1}^o \dots \mathcal{P}_0^o\}$, $\mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}}$ is given by Equation (18), and the initial value is set to its unconditional mean.

As discussed in Section 2.5, the canonical form $\Theta^{\mathcal{P}}$ is defined in terms of free factor loadings H. When imposing the no-arbitrage restriction, we estimate parameters of the canonical form $\widetilde{\Theta}^{\mathcal{P}}$ via the following two-step procedure. The first step corresponds exactly to maximizing the likelihood in Equation (32) for the whole sample. In the second step, we estimate the risk-neutral parameters $K_1^{\mathbb{Q}}$, δ_0 and δ_1, keeping the parameters $K_1^{\mathbb{P}}$, R and θ at values estimated in the first step. The likelihood is then given by

$$f(\mathcal{P}_t^o|\mathcal{I}_{t-1};\widetilde{\Theta}^{\mathcal{P}}) = -\frac{1}{2}\left(\log \Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}' + (\mathcal{P}_t^o - \mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}})'(\Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}')^{-1}(\mathcal{P}_t^o - \mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}})\right), \quad (33)$$

where the plugged-in values from the first stage are consistent estimates for the parameters of interest. Therefore, the second-stage estimates of $K_1^{\mathbb{Q}}$, δ_0 and δ_1 are consistent as well.[18] In all cases, we restrict the eigenvalues of $K_1^{\mathbb{Q}}$ and $K_1^{\mathbb{P}}$ within the unit circle, we calibrate the level of the risk factor under $\mathbb{P}$ to match the sample averages of $f_t^{(12)}$, $f_t^{(36)}$ and $f_t^{(60)}$, and we also calibrate δ_0 to match the average of the short rate $y_t^{(1)}$ in our sample. This makes the different models more easily comparable.

In cases where the risk factors can be inverted to reveal the risk factors, the parameterization reduces to the canonical form in Joslin, Singleton, and Zhu (2011). We also use their two-step estimation procedure. Denote $\overline{\mathcal{P}}_t = W\mathcal{P}_t^o$ the $\mathcal{N}$ yield portfolios that are measured without errors and $\mathcal{P}_t^e$ the portfolios that are measured with errors and redefine $\mathcal{P}_t^o = (\overline{\mathcal{P}}_t \;\; \mathcal{P}_t^e)$. Suppose that the measurement errors $\mathcal{P}^o - \mathcal{P}_t$ have conditional distribution P^σ, for some parameters σ, and that these errors are independent of lagged measurement errors.[19] In this case, the joint likelihood of $\mathcal{P}_t^o$ of all yields observed at time t is given by

$$f(\mathcal{P}_t^e|\overline{\mathcal{P}}_t; \lambda^{\mathbb{Q}}, k_\infty^{\mathbb{Q}}, \Sigma, P^\sigma) f(\overline{\mathcal{P}}_t|\mathcal{I}_{t-1}; K_0^{\mathbb{P}}, K_1^{\mathbb{P}}, \Sigma). \tag{34}$$

Parameters of the $\mathbb{P}$-dynamics are estimated in a first stage via equation-by-equation OLS regressions of $\overline{\mathcal{P}}_t$ on its lagged value $\overline{\mathcal{P}}_{t-1}$. Parameters of the $\mathbb{Q}$-dynamics are estimated in a second stage by minimizing the squared pricing errors for the remaining portfolios $\mathcal{P}_t^e$.

3.4 Markovian Models

The first question of interest is whether the non-Markovian dynamics are necessary to match the risk-premium evidence. For this purpose, we focus on CM_5^0-$KF0$ and CM_5^5-$KF0$ models with five factors. We endow the Markovian model with as many factors as forward rates in $\mathcal{P}_t$, so that the results cannot be attributed to information in the cross-section of yields being missed by a low-dimensional model. We drop the $KF0$ label in this case since R and θ are not separately identified (we estimate θ). We also use the 3-month and 6-month yields $y_t^{(3)}$ and $y_t^{(6)}$ to estimate the $\mathbb{Q}$-dynamics (we need at least one additional yield for $\mathcal{J} < \mathcal{N}$ and to identify the $\mathbb{Q}$-parameters). We will ask whether we can reduce the number of factors in later sections.

[18]The two-step estimator will be strongly consistent but not efficient for the parameters of the conditional mean dynamics in the $CM_{\mathcal{N}}$ models. However, one could construct a three-step Aitken type estimator that reaches the efficiency bound along the way suggested in Gallant (1975).

[19]The measurement errors also satisfy the consistency condition $\mathbb{P}(W\mathcal{P}_t^o = \overline{\mathcal{P}}_t) = 1$, as in Joslin et al. (2011).

3.4.1 Variance Ratios

Following Duffee (2011), we compare the variability of the bond risk premia from each model. This provides one measure of the gap between the information content of each model's state vector. Specifically, we compute the conditional risk premium for bonds with 6, 12, 24 and 60 months for holding periods between 1 and 12 months at each date in the sample. We then compute the sample variance for each case and within each model. Panel (A) of Table 6 reports the ratio of the sample variance from the CM_5^0 model relative to the CM_5^5. A value of less than one indicates that the CM_5^0 risk premium is less variable than the CM_5^5 risk premium, and gauges the additional information contained in the history yields. This contribution is large. The ratios typically hover around 40-50%. For instance, the ratio increases from 32% to 55% and from 18% to 59% for the 2-year and the 5-year bonds, respectively. The ratio for the 6-month and the 1-year bonds is also low at short horizons – 52% and 41%, respectively – but rises rapidly when we reach horizons near the bond maturity. Consistent with Duffee (2011) and the evidence from direct regressions in Table 1, the contribution from past yields increases at shorter horizons.

3.4.2 Bond Returns

In a second step, we ask whether the additional information content translates to a greater accuracy in forecasting actual excess returns. Specifically, we estimate the following Mincer-Zarnowitz regressions of excess returns on the model bond risk premium:

$$xr_{t,h}^{(n)} = a + b \times \widehat{brp}_{t,h}^{(n)} + u_{t,h}^{(n)}. \tag{35}$$

Panel (B) of Table 6 reports the ratios of the R^2s using the CM_5^0 risk premium relative to the CM_5^5 risk premium. A value of less than one gauges the accuracy gap in forecasting bond returns. This gap is large. In fact, the ratios in Panel (B) are close to the ratios in Panel (A): much of the added variability in the risk premium translates into an improved accuracy. The relative improvements are striking at the shorter horizons (in part because the predictability implied by the CM_5^0 is low), but remain large across the board. For instance, the R^2 ratios range between 13% and 72% for a 2-year bond, and between 27% and 64% for a 5-year bond.[20]

[20]Strictly speaking, the GDTSMs imply that $a = 0$ and $b = 1$ in Equation (35). Unreported results show that, with a few exceptions, this constraint yields little change in predictability.

3.4.3 Return-Forecasting Factors

How much of the variation in the return-forecasting factor in Section 1 is captured in the model-implied risk premium? Consider a regression of the factor estimated using Equation (3) on the risk premium implied by the models,

$$\hat{xr}^{(n)}_{t,h} = a + b \times \widehat{brp}^{(n)}_{t,h} + u^{(n)}_{t,h}. \tag{36}$$

Again, we report the ratio of the R^2s obtained from the CM_5^0 model relative to that obtained from the CM_5^5 model.[21] Figure 1 shows this ratio for bonds with 2, 3, 4 and 5 years to maturity and across horizons up to 12 months. Again, the improvements are substantial. The fit in the CM_5^0 model ranges between 55% and 85% of the fit in the CM_5^5 model for 3-month returns, and between 65% and 95% for 12-month returns.

3.4.4 CP's Regression

Relaxing the Markov assumption improves model forecasts even when keeping the information content fixed. Consider CP's regression of annual bond excess returns on $\mathcal{P}_t$. This keeps the information set constant. We then compare the population coefficients of this projection within the CM_5^0 and the CM_5^5 models, respectively. Otherwise, the information set used to form a forecast is larger in the CM_5^5 model than in the CM_0^5 model (the former uses the history of forward rates as well). This comparison is also analogous to using coefficients from the Campbell-Shiller regression as a benchmark to gauge DTSMs. Figure 2 shows the coefficients for bonds with maturities of 3 and 5 years and for horizons of 3, 6 and 12 months across Panels (A)-(C), including estimates from direct regressions. The OLS coefficients show the familiar tent shape. The coefficients from the CM_5^5 model are close to their sample counterpart and show the expected tent shape. But the coefficients from the CM_5^0 do not. CP also show that the CM_5^0 does not match these coefficients for annual returns. The fact that the CM_5^5 model provides a better fit means that it bridges the aggregation gap between the monthly yield dynamics and bond returns at longer horizons.

[21]Note that the variable on the left-hand side is measured with errors and the estimates of b may be biased. In addition, $\hat{xr}^{(n)}_{t,h}$ has been obtained from a finite sample and may suffer from overfitting. Hence, it is not clear how high successful R^2s should be in Equation (36).

3.5 Conditional Mean Models

This section focuses on models with $\mathcal{N} = 3$ factors and investigates the roots of the non-Markovian bond risk premium. CM models with $\mathcal{N} = 3$ are consistent with the stylized fact that three principal components summarize the cross-section of yields (i.e., $rank(var(\mathcal{P}_t)) = \mathcal{N} = 3$). These models are also consistent with the predictive content of lagged yields.

We first ask whether measurement errors on their own can match the evidence. To answer this question, we consider two models where Z_t is Markovian but where all yields are measured with errors: the CM_3^0-$KF1$ and CM_3^0-$KF2$. We can then assess how much information is hidden within yield measurement errors, as in Duffee (2011). If not, we ask whether the addition of non-Markovian risk factors matches the evidence. To answer this question, we consider two models where Z_t has non-Markovian dynamics – CM_3^3-$KF1$ and CM_3^3-$KF2$ – allowing us to gauge the effect of conditional mean dynamics on the estimated risk premium. In most cases, we compare results with the standard CM_3^0-$KF0$ model, where the risk factors are Markovian and can be inverted from three combinations of yields that are measured without error.

3.5.1 Likelihood Ratio Tests

Panel (A) of Table 7 reports the number of parameters in each model. The CM_3^0-$KF0$ has 24 parameters, including the standard deviation parameters for those combinations of yields measured with errors. The CM_3^0-$KF1$ has 23 parameters: all yields are measured with errors but only one parameter controls their variance, $R = \sigma^2 I$. The CM_3^0-$KF2$ has 27 parameters, since in this case the diagonal matrix R is free. Each corresponding model with non-Markovian risk factors has nine additional parameters (θ has $3 \times 3 = 9$ parameters). The CM_3^3-$KF0$, CM_3^3-$KF1$ and CM_3^3-$KF2$ models have 33, 32 and 36 parameters, respectively. To highlight the gain in parsimony, the CM_5^0-$KF0$ and CM_5^5-$KF0$ 5-factor models estimated in Section 3.4 have 51 and 76 parameters, respectively.

Panel (A) reports the gain in likelihood relative to CM_3^0-$KF1$. Note that the CM_3^0-$KF0$ model is estimated in two steps where non-linear least squares is applied in the second steps. Hence, we do not report the likelihood for this case.[22] The (log) likelihood increases by 85 between the CM_3^0-$KF1$ and CM_3^0-$KF2$. Similarly, the (log)

[22] We could compute the joint likelihood of the data at the two-step parameters estimates, but this would not correspond to the maximum likelihood point in the space of parameters (in our finite sample), and would not be comparable to the other results.

likelihood increases by 71 between the CM_3^3-$KF1$ and CM_3^3-$KF2$. In each case, the p-value of the LR test statistics associated with the null hypothesis that $R = \sigma^2 I$ is negligible (against the alternative that R is a free diagonal matrix). The restriction that the scale of measurement errors is driven by one parameter is rejected in the data. Turning to the restriction of Markovian risk factors, the likelihood increases by 41 between the CM_3^0-$KF1$ and CM_3^3-$KF1$ and by 27 between the CM_3^0-$KF2$ and CM_3^3-$KF2$ models (holding the form of the R matrix constant in each case). The restriction that $\theta = 0$ is rejected in both cases.[23]

3.5.2 Variance Ratios

Notwithstanding the statistical evidence, the question of interest is whether a 3-factor CM model captures the bond risk premium that we see in the data. We focus on annual bond returns in the following to preserve space. Results based on annual returns are also more easily comparable with the existing evidence. In addition, using annual returns yields a conservative assessment, since the relative importance of the non-Markovian effects increases for shorter horizons.

As above, we first compare the standard deviation of the model-implied risk premium. We compute the ratio relative to the unrestricted 5-factor CM_5^5 model, which is the relevant benchmark, since it captures the predictability observed in the data (see, e.g., Table 5). Panel (B) of Table 7 reports the ratios. A ratio close to or greater than one is favorable. First, the risk-premium variability is low in the CM_3^0-$KF0$, which corresponds to the standard 3-factor model (as in Joslin, Singleton, and Zhu (2011), say). The ratio is only 70%. Worse, the annual bond returns predictability is only 60% of that observed in the data. These results provide a gauge to assess the gain across 3-factor models.

The CM_3^0-$KF1$ model produces more variability in the risk premium, around 90% of the sample variability, but the returns predictability R^2s do not match the evidence (see Section 3.5.3). Allowing for a more flexible structure of measurement errors worsens the results. The CM_3^0-$KF2$ model sees the variance ratio decrease by 15%. Filtering the risk factors is not enough. In contrast, allowing for three non-Markovian risk factors matches the evidence in the data. Both the CM_3^3-$KF1$ and CM_3^3-$KF2$ produce ratios above one.

[23] A likelihood gain of 11.8 or more is enough to reject the null using a 1% level for the test.

3.5.3 Bond Returns Forecasts

We also compare the R^2s from Mincer-Zarnowitz regressions of annual returns. We focus on annual returns for parsimony, and compute R^2 ratios relative to the unrestricted 5-factor CM_5^5 model. Panel (C) of Table 7 reports the results. Some of the variability of the risk premium in Markovian models does not translate into returns predictability. The CM_3^0-$KF1$ model produces variance ratios of around 90% but the predictability R^2 ratios decline to between 75-85%. In contrast, the R^2 ratios remain close to one in the non-Markovian models.

3.5.4 Return-Forecasting Factors

We also ask how much is lost in fitting the returns factor when $\mathcal{N} = 3$. Consider a regression of excess returns on the model risk premium,

$$\hat{xr}_{t,h}^{(n)} = a + b \times \widehat{brp}_{t,h}^{(n)} + u_{t,h}^{(n)}. \tag{37}$$

Figure 3 shows the ratio of the R^2s from the CM_3^0-$KF2$ model relative to that obtained from the CM_3^3-$KF2$ model. This gauges the importance of allowing for non-Markovian risk factors. The ratios are close to 100% for 1-month returns, most likely because each model's estimates follow from maximizing the likelihood of 1-month forecast errors. More importantly, the ratios immediately drop to between 60% and 70% for 2-month returns, and the improvement offered by CM models declines only slowly for longer horizons.

3.6 Summary

We provide a brief summary of this section's results. First, the Markov assumption is restrictive even in a 5-factor model estimated on five yields. The risk premium estimated based on CM dynamics exhibits more variability, with higher correlation with other return-forecasting factors, and delivers more accurate forecasts of bond returns. Second, a parsimonious 3-factor CM specification is sufficient to match the evidence. For instance, the CM_3^3-$KF2$ model adds nine parameters relative to a Markovian specification of the risk factor, but the R^2s increase by as much as 30% in the predictability regression of bond returns. In other words, while the models agree on the 1-month-ahead dynamics (which corresponds to the maximum likelihood criteria), the aggregation of forecasts over longer horizons differs substantially. Finally, measurement errors generate conditional mean dynamics in yields, but fail to match the evidence: one needs to introduce conditional mean dynamics in the latent factors.

4 Yield Decomposition

Estimates of the risk premium from the more flexible CM models are close to estimates from predictability regressions. In turn, the previous section has also shown that the CM models provide better forecasts of bond returns. In other words, the model decomposition of yields into an expectation component and a risk component is corroborated by indirect evidence. Since the risk premium differs from the standard models, it is important to examine more closely how the decomposition of long-term yields changes in our models. Figure 4 compares the decomposition of the 5-year yields from three different models. The CM_3^3-$KF2$ model generates non-Markovian effects via the risk factors and via the Kalman filter. The CM_3^3-$KF0$ generates non-Markovian effects via the risk factors only. The CM_3^0-$KF0$ corresponds to the standard Markovian model, with three combinations of yields measured perfectly.

Panel (A) shows the decomposition from the CM_3^0-$KF0$ model. Estimates of the term premium exhibit little cyclical variability. Therefore, most of the 5-year yield variations are attributed to changes in expectations. Panel (B) shows the decomposition from the CM_3^3-$KF2$ specification. The term premium is substantially more variable and cyclical. Non-Markovian effects attribute a much greater share of 5-year variations to the risk premium, producing smoother estimates of interest rate expectations. Figure 5 compares the term premium from each model, providing a better scale to assess the magnitude of the differences. Again, the simple Markovian model produces much less term premium variation. For instance, the estimates remain between 2% and 3% in the early 1980s, when the 5-year yield reached close to 15%. In contrast, the CM_3^3-$KF2$ term premium estimates range between 4% and 5%, almost twice as large as the benchmark case.

The cyclical differences are also immediately apparent. The larger cyclical variations imply that, early in recessions (e.g., 1990-1991 and 2001), the expectation component falls by less and it recovers faster as the economy approaches a turning point. In fact, the episode between 2001 and 2005 provides a powerful case in point. Using the information from past yields produces a large decline of term premium estimates early in the 2001 recessions. The downward adjustment in expectations is far smaller in the early phase of the recessions. The lower expectation estimates are also counter-intuitive. It is unlikely that yields were expected to remain as low and for so long as suggested by the non-Markovian model.

Strikingly, this difference has further repercussions as we enter the "conundrum" episode in 2004-2005. The risk premium from the CM_3^3-$KF2$ model increases steadily as the Federal Reserve raises short-term interest rates, but the risk premium from the CM_3^0-$KF0$ *falls* throughout 2004 and 2005. Hence, the decline in long-term yields in that period is wrongly attributed to a fall in expectations, an explanation rejected at the time by Chairman Greenspan.[24] Other implementations of DTSMs also attribute the conundrum to an increase in the risk premium, but this typically requires sophisticated bias-adjusted estimators, or auxiliary assumptions allowing researchers to use surveys of professional forecasters. Our results suggest that misspecification of the benchmark VAR model plays an important role: simply relaxing $\theta = 0$ correctly describes the evolution of the risk premium.

5 Conclusion

We revisit and extend the evidence for non-Markovian effects in the dynamics of yields put forward in Cochrane and Piazzesi (2005) and Duffee (2011). Estimates of the bond risk premium are much more variable and cyclical once we account for the information content from past yields. In turn, these risk-premium estimates produce forecasts of bond returns that are more accurate. As noted by Cochrane and Piazzesi (2005) and Duffee (2011), among others, the evidence is inconsistent with a standard Markovian specification of the GDTSM. This paper proposes a simple, parsimonious and intuitive reconciliation, where we identify two separate channels generating non-Markovian dynamics for yields (under the historical measure $\mathbb{P}$). First, the latent risk factor may include a moving-average component. Second, the application of the Kalman filter generates non-Markovian effects from the point of view of the econometrician.

Empirically, this approach captures much of the predictive content of past forward rates for bond excess returns. We also show that measurement errors participate in the observed non-Markovian dynamics, but that they cannot on their own generate the observed pattern of predictability. In practice, our approach generates estimates of the risk premium that are more cyclical and more accurate. In turn, this generates economically large differences in the decomposition of long-term yields.

We leave several avenues for future research. For instance, additional data, such as bid-ask spreads, could be brought into the model to help identify the effect of measurement errors from the effect of non-Markovian risk factors. Also, we do not ask why the latent risk factors are not Markovian: is it because the true state exhibits

[24] See, e.g., Backus and Wright (2007) for a contemporaneous discussion.

long-run dynamics, as in Bansal and Yaron (2000); because preferences exhibit habit, as in Campbell and Cochrane (1999); or because investors form expectations adaptively? In any case, our approach offers researchers and investors an interpretation of bond yields that is consistent with the dynamic properties of bond returns.

References

Ang, A., S. Dong, and M. Piazzesi (2007). No-arbitrage Taylor rules. Working Paper 13448, National Bureau of Economic Research.

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

Arnold, W. F. and A. J. Laub (1984). Generalized eigenproblem algorithms and software for algebraic Riccati equations. *Proceedings of the IEEE 72*(12), 1746–1754.

Backus, D. and J. Wright (2007). Cracking the conundrum. *Brookings Papers on Economic Activity* (1:2007), 293–328.

Bansal, R. and A. Yaron (2000). Risks for the long run: A potential resolution of asset pricing puzzles. Working Paper 8059, NBER.

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

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.

Cieslak, A. and P. Povala (2013). Expecting the Fed. Working paper, Northwestern University.

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

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

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

Fama, E. (1984). Term premiums in bond returns. *Journal of Financial Economics 13*, 529–546.

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

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007). ABCs (and Ds) of understanding VARs. *American Economic Review 97*(3), 1021–1026.

Feunou, B. and J.-S. Fontaine (2012). Forecasting inflation and the inflation risk premiums using nominal yields. Working Paper 2012-37, Bank of Canada.

Fiorentini, G. and E. Sentana (1998). Conditional means of time series processes and time series processes for conditional means. *International Economic Review 39*(4), 1101–1118.

Froot, K. (1989). New hope for the expectations hypothesis of the term structure of interest rates. *The Journal of Finance 44*(2), 283–305.

Gallant, A. (1975). Seemingly unrelated nonlinear regressions. *Journal of Econometrics 3*(1), 35–50.

Griliches, Z. (1967). Distributed lags: A survey. *Econometrica 35*(1), 16–49.

Gurkaynak, R., B. Sack, and J. Wright (2006). The U.S. Treasury curve: 1961 to present. Technical Report 2006-28, Federal Reserve Board.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.

Jardet, C., A. Monfort, and F. Pegoraro (2008). No-arbitrage near-cointegrated VAR(p) term structure models, term premia and GDP growth. Working paper, Banque de France.

Johannes, M., L. Lochstoer, and Y. Mou (2011). Learning about consumption dynamics. Working paper, Columbia Business School.

Joslin, S., A. Le, and K. J. Singleton (2013). Gaussian macro-finance term structure models with lags. *Journal of Financial Econometrics*. (forthcoming).

Joslin, S., M. Priebsch, and K. J. Singleton (2012). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance*. (forthcoming).

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

Koyck, L. M. (1954). *Distributed Lags and Investment Analysis*. North-Holland Publishing Co. Amsterdam.

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

Ludvigson, S. and S. Ng (2009). Macro factors in bond risk premia. *Review of Financial Studies 22*(12), 5027–5067.

Monfort, A. and F. Pegoraro (2007). Switching VARMA term structure models. *Journal of Financial Econometrics 5*(1), 105–153.

Piazzesi, M. and M. Schneider (2009). Trend and cycle in bond premia. Working paper, Stanford University.

Ravenna, F. (2007). Vector autoregressions and reduced-form representations of DSGE models. *Journal of Monetary Economics 54*(7), 2048–2064.

Shiller, R. (1979). The volatility of long-term interest rates and expectation models of the term structure. *The Journal of Political Economy 87*, 1190–1219.

Startz, R. (1982). Do forecast errors of term premia really make a difference between long and short rates? *Journal of Financial Economics 10*, 323–329.

A Appendix

A.1 Bond Returns Projections with a Moving-Average Term

The factors Y_t have VARMA(1,1) dynamics with a VAR(∞) representation given by

$$Y_{t+1} = \sum_{i=0}^{\infty} \Theta_i Y_{t-i} + u_{t+1}, \tag{38}$$

with $\Theta_i \equiv \theta^i(\phi - \theta)$. Guess that $E[Y_{t+h}|Y_t, \ldots,] \equiv E_t[Y_{t+h}] = C_h \sum_{i=0}^{\infty} \Theta_i Y_{t-i}$ for $h > 1$ with $C_1 = I$, the identity matrix. Then,

$$\begin{aligned} E_t[Y_{t+h}] = E_t[E_{t+1}[Y_{t+h}]] &= E_t[C_{h-1} \sum_{i=0}^{\infty} \Theta_i Y_{t+1-i}] \\ &= C_{h-1} (I + \theta) \sum_{i=0}^{\infty} \Theta_i Y_{t-i}, \end{aligned} \tag{39}$$

where $C_h = C_{h-1} (I + \theta)$. Hence, for any $R_{t+h} \equiv W_h Y_{t+h}$, we have that

$$E_t[R_{t+h}] = W_h C_h \sum_{i=0}^{\infty} \Theta_i Y_{t-i}. \tag{40}$$

Therefore, the projection coefficients of R_{t+h} on Y_{t-i} for $i = 0, \ldots, \infty$ are given by $W_h C_h \Theta_i = W_h C_h \theta^i(\phi - \theta)$, which has the rank of θ, where we have a separation between the dependence on n and h (implicit in W_h) and the decaying pattern due to θ^i.

A.2 Proofs of Propositions 1-3

Proposition 1

Take $Z_t \in \mathbb{R}^{\mathcal{N}}$ with stationary Gaussian distribution under $\mathbb{Q}$, possibly conditional on Z_t, its past, or some other factor $Z_{2,t}$. Suppose that the cross-section of n-period yields, $Y_t^{(n)}$ for $n \geq 1$, can be expressed as a linear function of Z_t,

$$y_t^{(n)} = A_n + B_n' Z_t. \tag{41}$$

In the absence of an arbitrage opportunity, we have that

$$y_t^{(n)} = -\frac{1}{n} \log \left(E_t^{\mathbb{Q}} \left[\exp \left(-\sum_{i=0}^{n} y_{t+i}^{(1)} \right) \right] \right),$$

which is consistent with Equation (41) only if Z_t has Markovian dynamics under $\mathbb{Q}$. We can write

$$\Delta Z_t = K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}} Z_{t-1} + \Sigma_Z \epsilon_{Z,t}^{\mathbb{Q}}, \tag{42}$$

where $\epsilon^{\mathbb{Q}}_{z,t}$ is a standard Gaussian innovation. The relationship between the loadings A_n and B_n and the $\mathbb{Q}$-dynamics for Z_t is standard and given by the no-arbitrage price of zero-coupon bonds,

$$\begin{aligned} D_{t,n} &= E^{\mathbb{Q}}_t[e^{-\sum_{i=0}^{n-1} r_{t+i}}] \\ &= e^{\mathcal{A}_n + \mathcal{B}'_n Z_t}, \end{aligned} \tag{43}$$

for $n > 1$, where $\mathcal{A}_n$ and $\mathcal{B}_n$ satisfy the first-difference equations:

$$\begin{aligned} \mathcal{A}_{n+1} - \mathcal{A}_n &= K^{\mathbb{Q}\prime}_0 \mathcal{B}_n + \frac{1}{2}\mathcal{B}'_n \Sigma_Z \Sigma'_Z \mathcal{B}_n - \rho_0 \\ \mathcal{B}_{n+1} - \mathcal{B}_n &= K^{\mathbb{Q}\prime}_1 \mathcal{B}_n - \rho_1, \end{aligned} \tag{44}$$

where the coefficients for yields are $A_n = -\mathcal{A}_n/n$ and $B_n = -\mathcal{B}_n/n$, and ρ_0 and ρ_1 are the loadings on the short rate:

$$r_t = \rho_0 + \rho_1 Z_t. \tag{45}$$

Proposition 2

Let $Z_t \in \mathbb{R}^{\mathcal{N}}$ with Markovian $\mathbb{Q}$-dynamics and consider the change of measure M_{t+1} given by

$$M_{t+1} \equiv \exp\left(-\frac{\lambda'_t \Sigma_Z \Sigma'_Z \lambda_t}{2} - \lambda'_t \Sigma_Z \epsilon^{\mathbb{P}}_{t+1}\right). \tag{46}$$

Define $\epsilon^{\mathbb{P}}_t$ and $\mathcal{E}^{\mathbb{P}}_{Z,t-1}$:

$$Z_t = \mathcal{E}^{\mathbb{P}}_{Z,t-1} + \Sigma_Z \epsilon^{\mathbb{P}}_t. \tag{47}$$

Then,

$$\begin{aligned} E^{\mathbb{Q}}_t[\exp(u'\epsilon^{\mathbb{P}}_t)] &= E_t[M_{t+1}\exp(u'\epsilon^{\mathbb{P}}_t)] \\ &= \exp(\frac{u'u}{2} - u'\Sigma'_Z \lambda_t), \end{aligned} \tag{48}$$

which implies that $\epsilon^{\mathbb{P}}_t = -\Sigma'_Z \lambda_t + \epsilon^{\mathbb{Q}}_t$ and, therefore, that

$$\mathcal{E}^{\mathbb{Q}}_t - \mathcal{E}^{\mathbb{P}}_t = -\Sigma_Z \Sigma_Z \lambda_t. \tag{49}$$

Assume that the prices of risk are given by

$$\lambda_t \equiv \left(\Sigma_Z \Sigma'_Z\right)^{-1}\left(\lambda_0 + \lambda_1 Z_t + \lambda_2 \mathcal{E}^{\mathbb{P}}_{Z,t-1}\right). \tag{50}$$

Then, Equation (13) follows from substituting for $\mathcal{E}^{\mathbb{Q}}_t$ and λ_t in Equation (49), where the unique mapping between parameters of the $\mathbb{P}-$ and $\mathbb{Q}-$ dynamics is given in Proposition 2.

A.3 Conditional Mean Representation of the Kalman Filter

Theorem 1

Define the matrix $H^{\perp}$ such that $H^{\perp}H' = I_{\mathcal{N}}$. This matrix exists but it is not unique (also $H(H^{\perp})^{\top} = I_{\mathcal{N}}$). Note that $\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t} = H_0 + H'\mathcal{E}^{\mathbb{P}}_t$. Then, the unobserved portfolios $\mathcal{P}_t$ have the following dynamics:

$$\begin{aligned}\mathcal{P}_{t+1} &= \mathcal{E}^{\mathbb{P}}_{\mathcal{P},t} + H'\Sigma\epsilon^{\mathbb{P}}_t \\ \mathcal{E}^{\mathbb{P}}_{\mathcal{P},t+1} &= H'K^{\mathbb{P}}_0 - K^{\mathbb{P}}_1 H_0 + H'(K^{\mathbb{P}}_1 + I_{\mathcal{N}})H^{\perp}\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t} + H'\Sigma_{\mathcal{E}}\epsilon^{\mathbb{P}}_t,\end{aligned} \tag{51}$$

and the observed portfolios are measured with errors,

$$\mathcal{P}^o_t = \mathcal{P}_t + w_t, \tag{52}$$

with i.i.d. $w_t \sim N(0,R)$. We use the VAR representation of $\mathcal{P}_t$ dynamics to analyze the optimal filter. Define $\mathcal{X}'_t \equiv (\mathcal{P}'_t\ \mathcal{E}^{\mathbb{P}'}_{\mathcal{P},t})$. Then, the extended state-space representation is given by

$$\begin{aligned}\mathcal{X}_{t+1} &= \mu_X + F_X\mathcal{X}_t + Q^{1/2}_X\epsilon^{\mathbb{P}}_t \\ \mathcal{P}^o_t &= H_{0X} + H'_X\mathcal{X}_t + w_t,\end{aligned} \tag{53}$$

where

$$F_X = \begin{bmatrix} 0 & I_{\mathcal{J}} \\ 0 & H'(K^{\mathbb{P}}_1 + I_{\mathcal{N}})H^{\perp} \end{bmatrix} \quad H_X = \begin{bmatrix} I_{\mathcal{J}} \\ 0 \end{bmatrix} \quad Q^{1/2}_X = \begin{bmatrix} H'\Sigma \\ H'\Sigma_{\mathcal{E}} \end{bmatrix}. \tag{54}$$

Then, the Kalman filter forecasts $\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t} = E^{\mathbb{P}}_t[\mathcal{P}_{t+1}]$, conditional on the observations $\{\mathcal{P}^o_1, \ldots, \mathcal{P}^o_t\}$ and some initial value $\mathcal{E}^{\mathbb{P}}_{\mathcal{P},0}$ is given by

$$\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t} = H'_X F_X \left\{E_{t-1}[\mathcal{X}_t] + P_{X,t|t-1}H_X(H'_X P_{X,t|t-1}H_X + R)^{-1}\left(\mathcal{P}^o_t - \mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1}\right)\right\} \tag{55}$$

where

$$P_{X,t+1|t} = F_X\left[P_{X,t|t-1} - P_{X,t|t-1}H_X(H'_X P_{X,t|t-1}H_X + R)^{-1}H'_X P_{X,t|t-1}\right]F'_X + Q_X, \tag{56}$$

and where we define

$$P_{X,t|t-1} \equiv \begin{bmatrix} P^{11}_{t|t-1} & \left(P^{21}_{t|t-1}\right)' \\ P^{21}_{t|t-1} & P^{22}_{t|t-1} \end{bmatrix}.$$

Tedious algebra leads to

$$\begin{aligned}\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t} &= (I_{\mathcal{J}} - K^{\mathbb{P}}_{1\mathcal{P}})H_0 + H^{\top}\mu + H'(K^{\mathbb{P}}_1 + I_{\mathcal{N}})H^{\perp}\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1} \\ &\quad + \left(H'(K^{\mathbb{P}}_1 + I_{\mathcal{N}})H^{\perp}\left(P^{11}_{t|t-1} - H'\Sigma\Sigma'H\right) + H'\Sigma_{\mathcal{E}}\Sigma'H\right)(P^{11}_{t|t-1} + R)^{-1}\left(\mathcal{P}^o_t - \mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1}\right).\end{aligned} \tag{57}$$

Standard results show that $P_{X,t|t-1}$ converges quickly with t, $P_{X,t|t-1} \to P_X$, where P_X solves

$$P_X = F_X[P_X - P_X H_X(H'_X P_X H_X + R)^{-1}H'_X P_X]F'_X + Q_X,$$

if a solution to the algebraic Ricatti equation exists (see Arnold and Laub 1984). Then, dropping the superscript on the upper left block, $P^{11}_{t|t-1} \to P^{11} \equiv P$, we can rewrite the steady-state version of Equation (57) as

$$
\begin{aligned}
\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t} &= (I_{\mathcal{J}} - K^{\mathbb{P}}_{1\mathcal{P}})H_0 + H'K^{\mathbb{P}}_0 + H'(K^{\mathbb{P}}_1 + I_{\mathcal{N}})H^{\perp}\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1} \\
&\quad + \Big(H'(K^{\mathbb{P}}_1 + I_{\mathcal{N}})H^{\perp}\left(P - H'\Sigma\Sigma'H\right) + H'\Sigma_{\mathcal{E}}\Sigma'H\Big)(P+R)^{-1}\left(\mathcal{P}^o_t - \mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1}\right) \\
&= K^{\mathbb{P}}_{0\mathcal{P}} + H'(K^{\mathbb{P}}_1 + I_{\mathcal{N}})H^{\perp}\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1} \\
&\quad + \Big(H'(K^{\mathbb{P}}_1 + I_{\mathcal{N}})H^{\perp}P - H'\theta\Sigma\Sigma'H\Big)(P+R)^{-1}\left(\mathcal{P}^o_t - \mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1}\right) \\
&= K^{\mathbb{P}}_{0\mathcal{P}} + (K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}})\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1} + \Big((K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}})P - H'\theta\Sigma\Sigma'H\Big)(\Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}})^{-1}\left(\mathcal{P}^o_t - \mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1}\right) \\
&= K^{\mathbb{P}}_{0\mathcal{P}} + (K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}})\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1} + \Sigma_{\mathcal{E}_{\mathcal{P}}}\epsilon^{\mathbb{P}}_{\mathcal{P},t}, \qquad (58)
\end{aligned}
$$

where the last three inequalities follow from the parameter definitions in the proposition as well as $\epsilon^{\mathbb{P}}_{\mathcal{P},t} = \Sigma^{-1}_{\mathcal{P}}\left(\mathcal{P}^o_t - \mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1}\right)$.

Figure 1: Fitting the Cochrane-Piazzesi Forecast in 5-factor Models

Ratios of R^2s from regressions of the returns forecasting factor – estimated from Equation (3) – on the model-implied bond risk premia from the CM_5^0 relative to the CM_5^5 model. Estimation based on a sample from January 1964 until December 2007. Returns computed from the GSW data set.

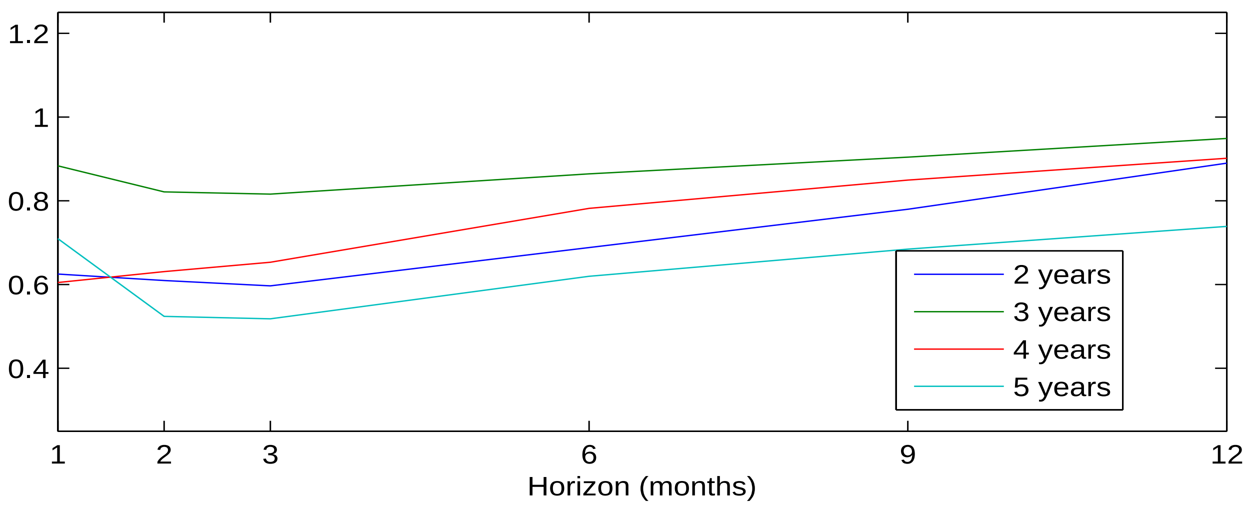
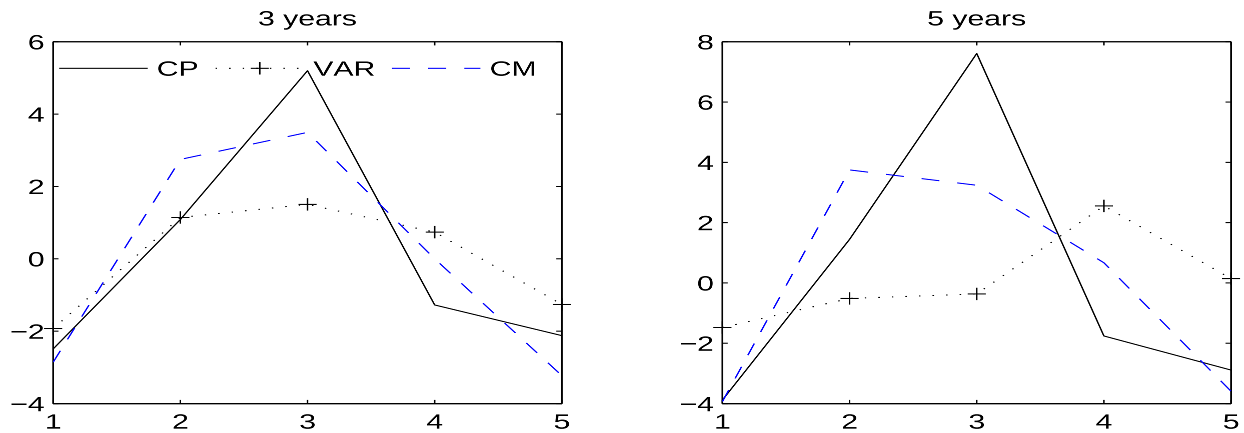


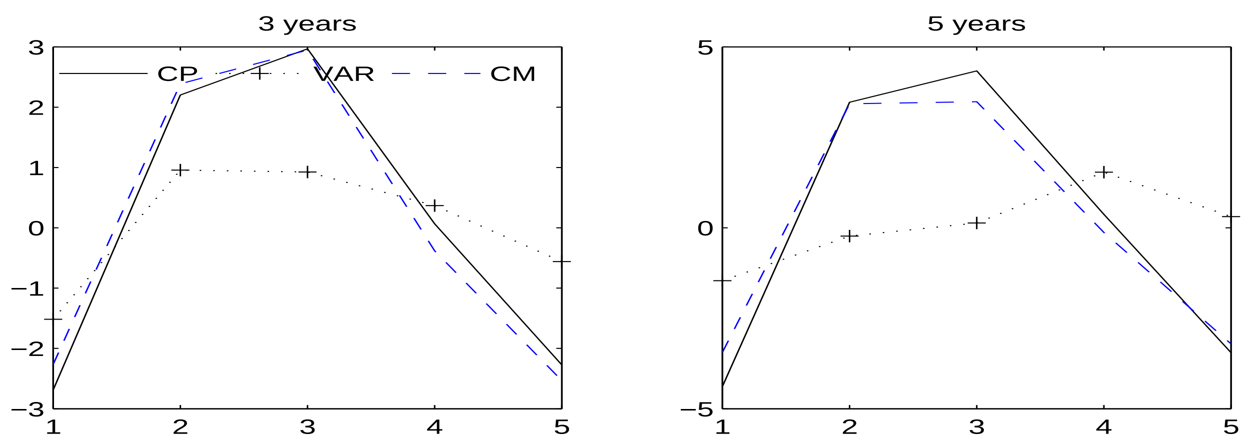
Figure 2: Predictability Coefficients in 5-Factor Models

Coefficients on forward rates from predictability regressions of bond excess returns on a constant and the forward rates $f_t = [1 f_t^{(1)}\ f_t^{(2)}\ f_t^{(3)}\ f_t^{(4)}\ f_t^{(5)}]'$. Each panel displays the OLS coefficient estimates, as well as the coefficients implied by the CM_5^0 and the CM_5^5 model, respectively. Panel (A) reports results for 3-month returns, Panel (B) reports results for 6-month returns, and Panel (C) reports results for 12-month returns. Each panel reports results for 2-year and 5-year bonds. Estimation based on a sample from January 1964 until December 2007. Returns computed from the GSW data set.

(A) 3-month returns



(B) 6-month returns



(C) 12-month returns

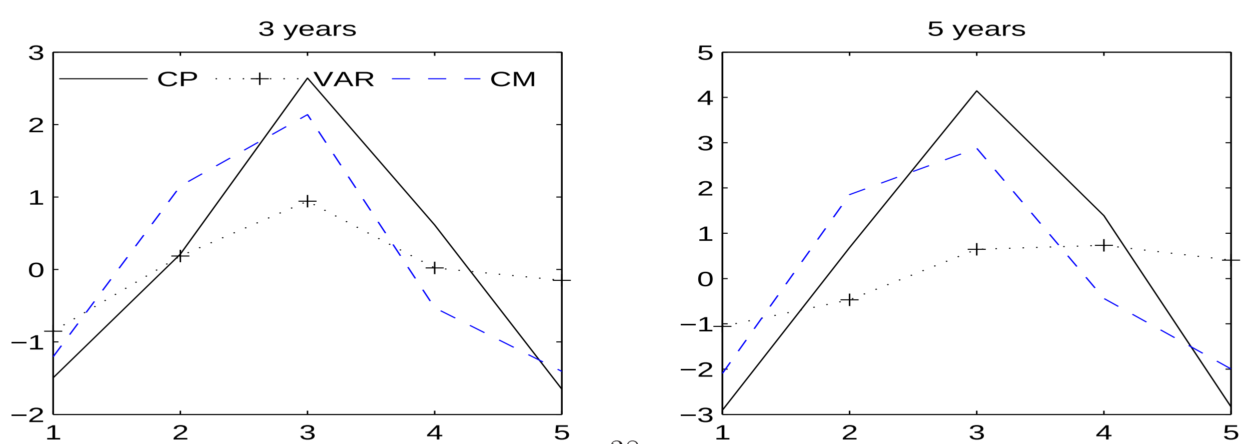


Figure 3: Fitting the Cochrane-Piazzesi Forecast in 3-factor Models

Ratios of R^2s from regressions of the returns forecasting factor – estimated from Equation (3) – on the model-implied bond risk premia from the CM_3^0-$KF2$ relative to the CM_3^3-$KF2$ model. Estimation based on a sample from January 1964 until December 2007. Returns computed from the GSW data set.

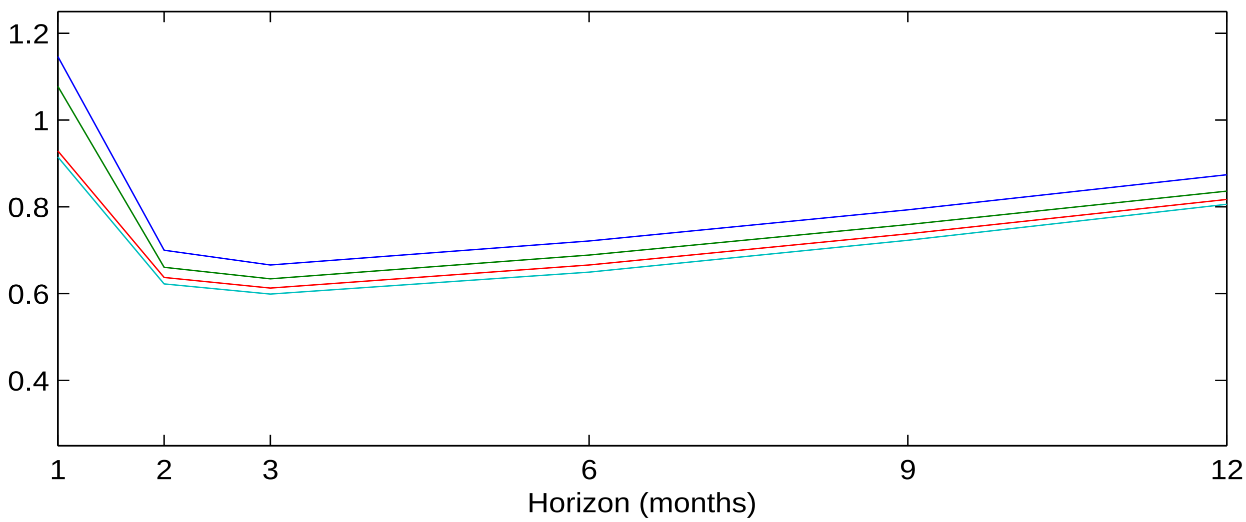
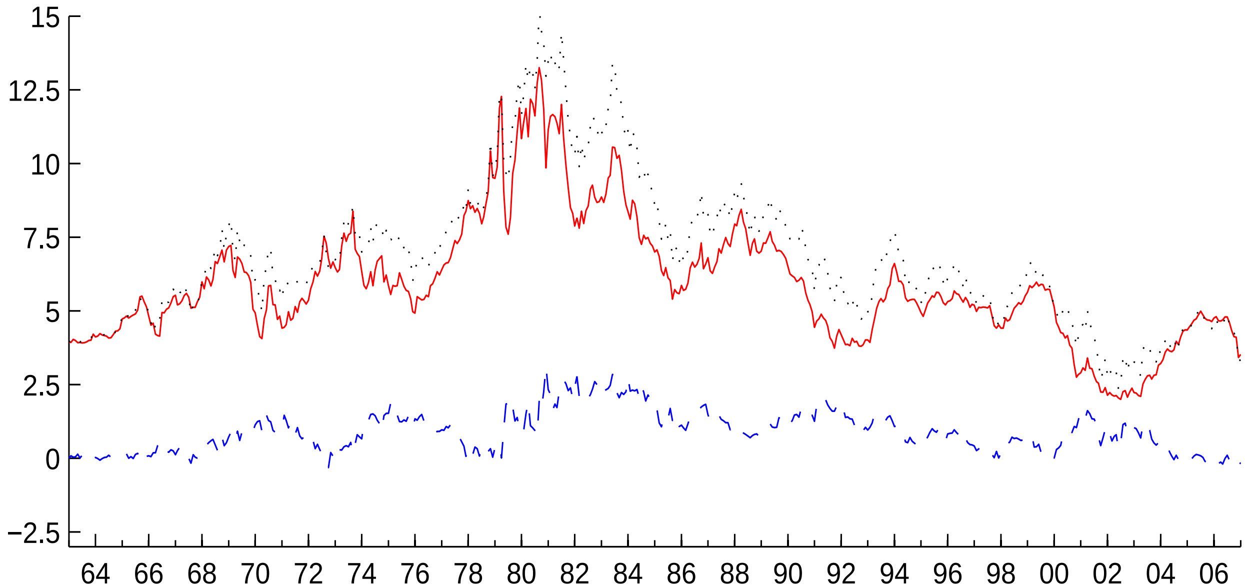


Figure 4: Yield Decompositions

Decomposition of the 5-year yield implied by three different models. The CM_3^3-$KF2$ has three non-Markovian risk factors and a free diagonal measurement error covariance matrix. The CM_3^0-$KF0$ has three Markovian risk factors and three forward rates measured without errors. Estimation based on a sample from January 1964 until December 2007.

(A) CM_3^0-$KF0$



(B) CM_3^3-$KF2$

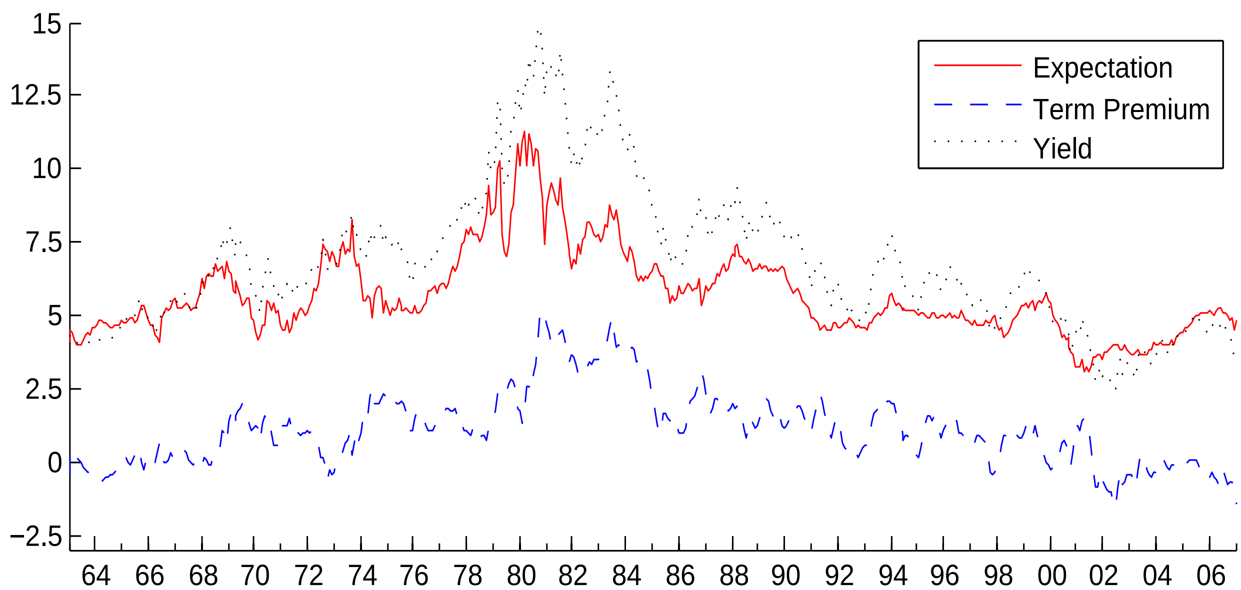


Figure 5: Term Premium

Term premium in 5-year yield implied by different models. The CM_3^3-$KF2$ has three non-Markovian risk factors and a free diagonal measurement error covariance matrix. The CM_3^0-$KF0$ has three Markovian risk factors and three forward rates measured without errors. Estimation based on a sample from January 1964 until December 2007.

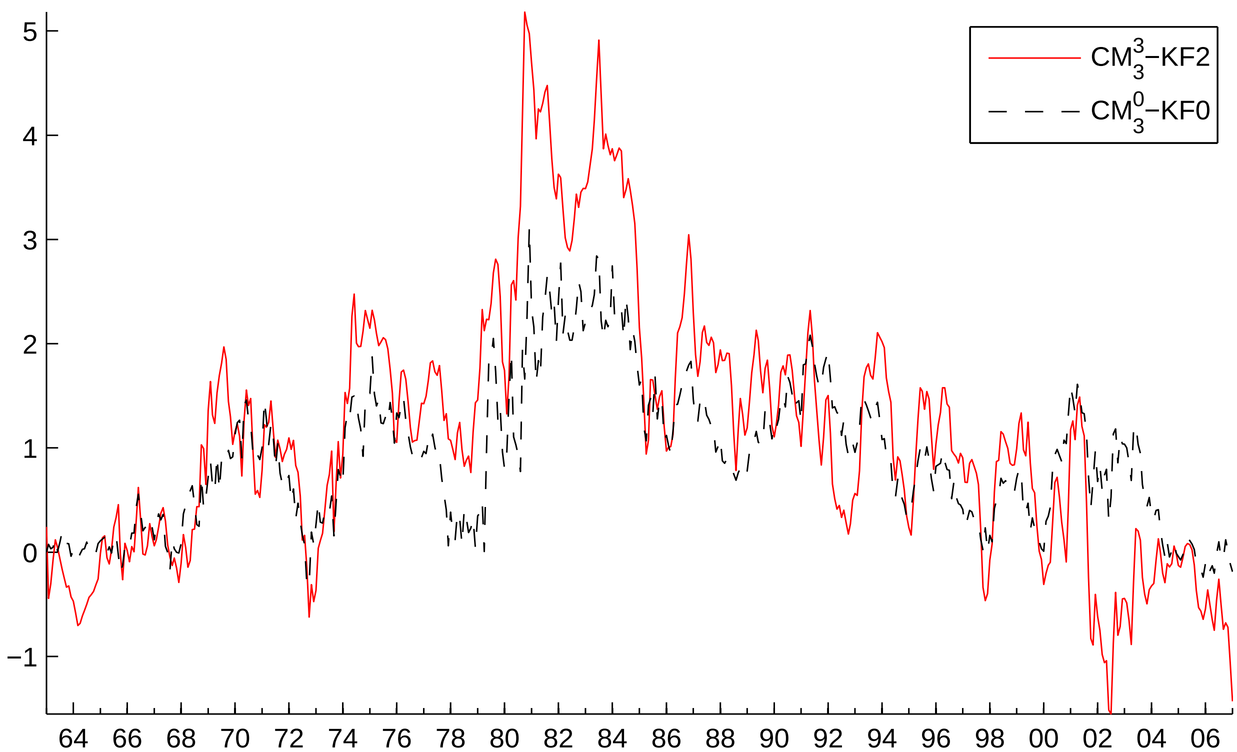


Table 1: **Bond Excess Returns – OLS Predictability Regressions**

Predictability of excess returns $xr_{t,h}^{(n)}$ on bonds with $n = 2, 3, 4, 5, 7, 10$ years to maturity for holding horizons of $n = 1, 2, 3, 6, 9, 12$ months using annual forward rates with 1, 2, 3, 4, and 5 years to maturity. Panel (A) reports R^2s from the predictability regressions,

$$xr_{t,h}^{(n)} = b_{n,h}\gamma_h' f_t + u_{t,h}^{(n)},$$

where f_t stacks a constant with annual forward rates with 1, 2, 3, 4, and 5 years to maturity, $b_{n,h}$ is a maturity-specific scalar and γ_h is an horizon-specific vector of coefficients. Panel B reports R^2s from the distributed-lag predictability regressions,

$$xr_{t,h}^{(n)} = b_{n,h}\left((1-\alpha_h)\gamma_h' f_t + \alpha_h \mathcal{R}_{t-1,h}\right) + u_{t,h}^{(n)},$$

where α_h is an horizon-specific scalar. Returns computed monthly from the GSW data set between Jan. 1963 and Dec. 2003.

Panel (A) CP Regressions

	Holding Period (months)					
Maturity	1	2	3	6	9	12
2	2.9	6.4	9.8	19.5	28.5	32.5
3	3.0	6.8	10.2	20.6	30.0	34.4
4	3.1	6.9	10.3	21.1	31.0	35.8
5	3.2	6.9	10.3	21.2	31.4	36.6
7	3.1	6.6	9.7	20.5	31.3	37.1
10	2.7	5.9	8.4	18.9	30.0	36.7

Panel (B) Recursive CP Regressions

	Holding Period (months)					
Maturity	1	2	3	6	9	12
2	6.6	10.4	15.1	26.8	35.9	35.7
3	7.0	11.3	16.8	30.1	39.4	40.0
4	7.2	11.9	18.0	32.6	42.0	43.0
5	7.3	12.2	18.7	34.4	43.8	45.0
7	7.1	12.3	19.2	36.3	45.8	47.3
10	6.5	11.7	18.6	36.7	46.7	48.6

Table 2: **Bond Excess Returns – Recursive Predictability Regressions**

Predictability of excess returns $xr^{(n)}_{t,h}$ on bonds with $n = 2, 3, 4, 5, 7, 10$ years to maturity for holding horizons of $n = 1, 2, 3, 6, 9, 12$ months using annual forward rates with 1, 2, 3, 4, and 5 years to maturity. Panel (A) reports estimates of α_h in the distributed-lag regression

$$xr^{(n)}_{t,h} = b_{n,h}\left((1-\alpha_h)\gamma'_h f_t + \alpha_h \mathcal{R}_{t-1,h}\right) + u^{(n)}_{t,h}, \tag{59}$$

where f_t stacks a constant with the forward rates. Panel (B) reports estimates of the scalars $b_{n,h}$ and Panel (C) reports estimates of the horizon-specific vector γ_h. Returns computed monthly from the GSW data set between Jan. 1963 and Dec. 2003.

Panel (A) α_h

	Holding Period (months)					
	1	2	3	6	9	12
	0.8320	0.8298	0.8170	0.7749	0.7096	0.6284

Panel (B) $b_{n,h}$

	Holding Period (months)					
Maturity	1	2	3	6	9	12
2	0.47	0.46	0.44	0.39	0.35	0.29
3	0.67	0.66	0.65	0.61	0.59	0.55
4	0.85	0.85	0.84	0.82	0.81	0.79
5	1.01	1.01	1.01	1.01	1.01	1.01
7	1.30	1.31	1.32	1.35	1.37	1.40
10	1.69	1.70	1.74	1.83	1.88	1.95

Panel (C) γ_h

	Holding Period (months)					
γ_h	1	2	3	6	9	12
γ_0	-0.04	-0.03	-0.03	-0.04	-0.04	-0.04
γ_1	-9.53	-9.40	-9.00	-6.90	-5.06	-4.03
γ_2	8.87	9.78	10.09	6.34	3.21	2.81
γ_3	11.84	9.51	6.85	6.67	6.23	3.73
γ_4	2.87	3.78	5.82	5.52	4.70	4.27
γ_5	-13.67	-13.42	-13.59	-11.32	-8.63	-6.32

Table 3: **Bond Excess Returns – Common Factor**

Principal-component analysis of the return-forecasting factors $\mathcal{R}_{t,h}$ in the distributed-lag regression

$$xr_{t,h}^{(n)} = b_{n,h}\left((1-\alpha_h)\gamma_h' f_t + \alpha_h \mathcal{R}_{t-1,h}\right) + u_{t,h}^{(n)} \tag{60}$$

estimated across horizons of 1, 2, 3, 6, 9, and 12 months.

	Principal Component					
	1	2	3	4	5	6
1	0.46	-0.42	-0.60	-0.08	0.28	-0.41
2	0.45	-0.36	-0.07	0.23	-0.34	0.71
3	0.45	-0.16	0.66	0.24	-0.20	-0.48
6	0.41	0.17	0.34	-0.42	0.64	0.31
9	0.36	0.47	-0.18	-0.54	-0.57	-0.10
12	0.30	0.65	-0.22	0.64	0.18	-0.004
R^2	96.8	2.7	0.31	0.11	0.02	0.003

Table 4: **Residuals from the VAR1 Model**

Panel (A) reports coefficient estimates $\hat{b}^{(n)}_{\epsilon,h}$ in regressions of each element of $\hat{\epsilon}_t$ on its own lag: $\hat{\epsilon}^{(n)}_t = b^{(n)}_{\epsilon,h}\ \hat{\epsilon}^{(n)}_{t-h} + w^{(n)}_{t,h}$, where $\hat{\epsilon}_t$ is the estimated residual from a Markovian model estimated for the yield vector Y_t, and where we vary $h = 1, \ldots, 6$. Panel (B) reports coefficient estimates $\hat{c}^{(m)}_{\epsilon,1}$ in regressions of each element of $\hat{u}_t$ on each element of the first lag of the residual $\hat{\epsilon}^{(n)}_t = c^{(m)}_{\epsilon,1}\ \hat{\epsilon}^{(m)}_{t-1} + w^{(n,m)}_{t,1}$. Panels (C) and (D) report coefficient estimates $\hat{\gamma}_\epsilon$ and $\hat{\alpha}_\epsilon$ from a two-step estimation of the single-factor regression $\hat{\epsilon}_t = \alpha_\epsilon \gamma_\epsilon^\top\ \hat{\epsilon}_{t-1} + w_t$, respectively, including the p-value of the F-statistic associated with the null hypothesis that γ_ϵ is zero. Asymptotic t-statistics shown in parentheses.

Panel (A) Own Lags

	$\epsilon^{(1)}_{Y,t}$	$\epsilon^{(2)}_{Y,t}$	$\epsilon^{(3)}_{Y,t}$	$\epsilon^{(4)}_{Y,t}$	$\epsilon^{(5)}_{Y,t}$
$\epsilon^{(n)}_{Y,t-1}$	0.15	0.17	0.12	0.09	0.08
	(1.77)	(2.25)	(2.06)	(1.47)	(1.31)
$\epsilon^{(n)}_{Y,t-3}$	-0.07	-0.07	-0.11	-0.10	-0.08
	(-0.72)	(-0.77)	(-1.25)	(-1.28)	(-1.31)
$\epsilon^{(n)}_{Y,t-5}$	-0.06	-0.08	-0.11	-0.05	-0.06
	(-0.82)	(-1.21)	(-1.80)	(-0.88)	(-1.00)

Panel (B) Cross-correlations

	$\epsilon^{(1)}_{Y,t}$	$\epsilon^{(2)}_{Y,t}$	$\epsilon^{(3)}_{Y,t}$	$\epsilon^{(4)}_{Y,t}$	$\epsilon^{(5)}_{Y,t}$
$\epsilon^{(1)}_{Y,t-1}$	0.15	0.14	0.09	0.06	0.06
	(1.77)	(2.00)	(1.50)	(1.15)	(1.18)
$\epsilon^{(2)}_{Y,t-1}$	0.18	0.17	0.11	0.08	0.08
	(2.06)	(2.25)	(1.72)	(1.36)	(1.44)
$\epsilon^{(3)}_{Y,t-1}$	0.20	0.19	0.12	0.10	0.09
	(2.44)	(2.72)	(2.06)	(1.65)	(1.83)
$\epsilon^{(4)}_{Y,t-1}$	0.20	0.19	0.12	0.09	0.09
	(2.19)	(2.43)	(1.83)	(1.47)	(1.61)
$\epsilon^{(5)}_{Y,t-1}$	0.19	0.18	0.12	0.08	0.08
	(1.94)	(2.20)	(1.65)	(1.21)	(1.31)

Panel (C) Single-factor restriction – first step

	$\epsilon^{(1)}_{Y,t-1}$	$\epsilon^{(2)}_{Y,t-1}$	$\epsilon^{(3)}_{Y,t-1}$	$\epsilon^{(4)}_{Y,t-1}$	$\epsilon^{(5)}_{Y,t-1}$	p-**val**
$\hat{\gamma}^{(n)}_\epsilon$	-0.08	0.04	0.37	0.17	-0.40	
	(-0.52)	(0.12)	(0.70)	(0.32)	(-0.84)	(0.02)

Panel (D) Single-factor restriction – second step

	$\epsilon^{(1)}_{Y,t}$	$\epsilon^{(2)}_{Y,t}$	$\epsilon^{(3)}_{Y,t}$	$\epsilon^{(4)}_{Y,t}$	$\epsilon^{(5)}_{Y,t}$
$\hat{\alpha}^{(n)}_\epsilon$	1.38	1.25	0.84	0.77	0.77
	(2.59)	(2.75)	(2.23)	(2.03)	(2.32)

Table 5: **Bond Excess Returns – Model-Implied Predictability**

Predictability of 1-year excess returns on bonds with 2, 3, 4, and 5 years to maturity where the predictors are the annual forward rates with 1, 2, 3, 4, and 5 years to maturity (Equation (1)). The table reports the predictability R^2s implied by a VAR(1), a VARMA(1,1) and a VAR(12) model for the vector of yields with 1, 2, 3, 4 and 5 years to maturity.

		Model	
Maturity	VAR(1)	VARMA(1,1)	VAR(12)
2	14.7	25.8	30.8
3	13.5	28.1	33.0
4	15.1	32.8	35.7
5	14.6	32.4	33.1

Table 6: **Sample Bond Risk Premium**

Ratios of sample bond risk-premium volatility and ratios of sample predictability R^2s from the CM_5^0 and CM_5^5 models. For each model and date, we compute the conditional bond risk premium in Equation (29) for bonds with 6, 12, 24 and 60 months to maturity and for holding periods between 1 and 12 months. Panel (A) reports the ratios of the sample variance in the CM_5^0 relative to that in the CM_5^5. Panel (B) reports the ratios of R^2s from Mincer-Zarnowitz regressions of bond excess returns on each model-implied risk premium. Estimation based on CRSP data from January 1964 until December 2007. Observed returns computed from the GSW data set.

Panel (A) Variance ratios

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
6	0.52	0.68	0.88	0.87	0.88							
12	0.41	0.62	0.68	0.72	0.74	0.76	0.78	0.79	0.80	0.81	0.82	
24	0.32	0.44	0.48	0.48	0.49	0.49	0.50	0.51	0.52	0.53	0.54	0.55
60	0.18	0.31	0.36	0.38	0.40	0.42	0.45	0.47	0.50	0.53	0.56	0.59

Panel (B) R^2 ratios

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
6	0.53	0.53	0.54	0.64	0.80							
12	0.13	0.22	0.21	0.36	0.45	0.48	0.56	0.58	0.61	0.67	0.76	
24	0.13	0.26	0.30	0.47	0.55	0.56	0.61	0.63	0.64	0.67	0.69	0.72
60	0.27	0.33	0.32	0.40	0.45	0.47	0.51	0.53	0.55	0.58	0.61	0.64

Table 7: **Conditional Mean Models with Three Factors**

Results from $CM_{\mathcal{N}}^{k} - KF$ models with $\mathcal{N} = 3$. The first line shows the label of each model. Panel (A) reports the number of parameters for each model, and the gain in likelihood relative to the most restricted case $CM_3^0 - KF1$. Panel (B) reports the ratio of the sample standard deviation of each model-implied bond risk premium relative to the 5-factor CM_5^5 model (see Section 3.4) for bond maturities of 24, 36, 48 and 60 months and for a holding period of 12 months. Panel (C) reports the ratio of R^2s in Mincer-Zarnowitz regressions of realized returns on the model risk premium. Estimation based on CRSP data from January 1964 until December 2007. Realized returns computed from the CRSP data.

	$CM_3^0 - KF0$	$CM_3^0 - KF1$	$CM_3^3 - KF1$	$CM_3^0 - KF2$	$CM_3^3 - KF2$
			Panel (A) Likelihood gains and parameters		
	24	23	32	27	36
		0	41	85	112
			Panel (B) Variance ratios		
2	0.73	0.97	1.07	0.84	1.05
3	0.71	0.91	1.03	0.77	1.04
4	0.77	0.90	1.04	0.76	1.04
5	1.02	0.92	1.04	0.79	1.05
			Panel (C) R^2 ratios		
2	0.66	0.80	1.02	0.70	0.92
3	0.64	0.81	1.03	0.70	0.97
4	0.63	0.84	1.02	0.75	1.01
5	0.48	0.75	1.00	0.63	1.03