Bond Risk Premia and Gaussian Term Structure Models

Bruno Feunou   Jean-Sébastien Fontaine
Bank of Canada

December 2015

Abstract

Cochrane and Piazzesi (2005) show that (i) lagged forward rates help predict bond returns and that (ii) modern Markovian dynamic term structure models (DTSMs) cannot match the evidence. We develop the family of Conditional Mean DTSMs where the dynamics depend on current yields and their history through a moving-average component. Our preferred Conditional Mean model combines one moving-average with the usual three Gaussian risk factors, closely matches the bond risk premium measured from predictive regressions and provides better forecasts of bond returns. Our framework nests Duffee (2011) models with a small "hidden" factor and our results compare favorably with his 5-factor model. Conditional Mean models are easier to estimate than state-space term structure models based on Kalman estimates of latent factors.

Keywords: Term Structure Models, Bond Risk Premium, Unspanned Risk

JEL Classification: E43, E47, G12

We thank Greg Bauer, Antonio Diez de los Rios, Greg Duffee, Anh Le, Dora Xia and participants at the 2013 Bank of Canada Fixed Income conference for comments and suggestions. We also the editor and two anonymous referees for constructive comments; and Glen Keenleyside for editorial assistance. Email: bfeunou@bankofcanada.ca or jsfontaine@bankofcanada.ca.

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, 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", documenting that past forward rates help forecast bond returns.[1] 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-DTSMs), 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, intuitive, and grounded in empirical evidence. Estimation is straightforward. The implied risk-premium variations are sizably larger and more cyclical than in standard specifications.

To understand the underlying tension, 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 $\mathcal{J} > \mathcal{N}$ yield portfolios $\mathcal{P}_t^o$ is connected linearly to $\mathcal{N}$ risk factors Z_t. Next, consider the standard assumption that risk factors Z_t are Markovian under the historical distribution $\mathbb{P}$. This is a time-serie statement: the best forecasts of $\mathcal{P}_{t+h}^o$ are also connected linearly to Z_t. Together, these assumptions imply that our best forecasts of future yields $\mathcal{P}_{t+h}^o$ (and bond returns) should be given by a linear combination of current yields $\mathcal{P}_t^o$. This implication is rejected in the data.

Hence, the standard framework ties up (i) the risk factors required to match the cross-section of yields and (ii) the information set $\mathcal{I}_t$ required to generate optional forecasts, given by the conditional expectations $E^{\mathbb{P}}[\mathcal{P}_{t+h}^o|\mathcal{I}_t]$. But there is no necessary connection between the two. Regarding point (i), we make clear in the following that if yields are linear in $\mathcal{N}$ conditionally Gaussian risk factors Z_t, then no-arbitrage requires that Z_t have Markovian

[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."

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. Point (ii) is the motivation for our paper. We break away from using $\{\mathcal{P}_t^o\}$ for the information set $\mathcal{I}_t$, and we expand this set to $\mathcal{I}_t = \mathcal{I}_{\mathcal{P},t} \equiv \{\mathcal{P}_t^o\ , \mathcal{P}_{t-1}^o\ , \ldots\}$, the history of yields.

The first part of the paper provides stylized predictability and time-series facts. We study the empirical properties of $E[\mathcal{P}_{t+h}^o|\mathcal{I}_{\mathcal{P},t}]$. We revisit the predictability of bond returns for investment 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 adding several lags of the return-forecasting factors add to the predictability evidence. The relative contribution of past forward rates increases for shorter horizons and peaks at the monthly horizon. Third, the coefficients on additional lags of the returns-forecasting factor exhibit a clear decaying pattern. We show analytically that the evidence is consistent with Gaussian models where the conditional mean $E[\mathcal{P}_{t+1}^o|\mathcal{I}_{\mathcal{P},t}]$ is Markovian but where $\mathcal{P}_t^o$ is not Markovian. Finally, we find that the information contained in the history of yields exhibits a tight factor structure, with possibly only one underlying predictive factor, which is consistent with results in Cochrane and Piazzesi (2005).

The second part of the paper introduces the class of CM-DTSMs. Consistent with the cross-section evidence, the $\mathcal{N}$ latent risk factors Z_t are Markovian under $\mathbb{Q}$. Consistent with the predictability evidence, the process for the conditional mean $\mathcal{E}_{Z,t} \equiv E[Z_{t+1}|\mathcal{I}_{Z,t}]$ under $\mathbb{P}$ depends on the history $\mathcal{I}_{Z,t} \equiv \{Z_t, Z_{t-1}, \ldots\}$. To get a sense of its properties, this process is analogous to GARCH dynamics where the (squared) innovations drive the factors' conditional variance. In a Conditional Mean models, the $\mathcal{N}$ innovations $u_t \equiv Z_t - \mathcal{E}_{Z,t-1}$ drive the process for $\mathcal{E}_{Z,t}$. This is parsimonious: there is no need to introduce more sources of risk (more shocks). More important, using the same shocks generates a separation between the *spanning* role of Z_t and the *forecasting* role of $\mathcal{E}_{Z,t}$. Hence, CM-DTSMs have $2 \times \mathcal{N}$

state variables, but with only $\mathcal{N}$ shocks.[2] Our approach collapses to the standard GDTSM framework if Z_t is Markovian (i.e., $\mathcal{E}_{Z,t} = \Phi Z_t \quad \forall t$).

CM-DTSM models identify two distinct mechanisms behind the information content of past yields. The first mechanism is due to measurement errors and to the limited information set of the econometrician, as in Duffee (2011). This mechanism exists even if Z_t is Markovian, in which case the observed yields have CM dynamics with several restrictions due to the structure of the measurement error covariance matrix R. To see this, suppose every yield portfolio $\mathcal{P}_t^o$ is measured with errors. Then, the information set available to the econometrician is $\mathcal{I}_{\mathcal{P},t}$. Starting from a Markovian DTSM but conditioning on $\mathcal{I}_{\mathcal{P},t}$ (and not $\mathcal{I}_{Z,t}$), we show how to construct an equivalent CM-DTSM for $\mathcal{P}_t^o$, where the conditional mean of the porfolios $\mathcal{E}_{\mathcal{P},t} \equiv E[\mathcal{P}_{t+1}^o | \mathcal{I}_{\mathcal{P},t}]$ is Markovian. The second mechanism exists when the risk factors themselves have CM dynamics. This mechanism exists even if yields are observed without errors. It may seem that these two mechanisms cannot be identified separately. By contrast, our Proposition 2 shows that we can identify both mechanisms under reasonable conditions, allowing us to measure to distinguish the roles of measurement errors and that of a moving-average component.

The third part of the paper makes three empirical points, showing that the data strongly favor CM-DTSMs. First, the Markov restriction for Z_t is soundly rejected in every case we consider. Even when allowing for $\mathcal{N} = 5$ risk factors, as in Duffee (2011). Adrian, Crump, and Moench (2013) also emphasize a specification using five principal components of yields as risk factor. Using five factors gives most flexibility for the standard model to capture the information from past and current yields. Yet, we find that a CM-DTSM model produces better forecasts of excess returns. Models where Z_t is Markovian produce half of the risk premium variability and much lower predictive R^2s relative to a CM-DTSM specification. The deterioration is worse for shorter returns horizons; including at the monthly horizon. In addition, CM-DTSMs mostly match forward rate coefficients estimated from Cochrane-Piazzesi regressions.

[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, which collapses to the standard VAR(1) specification when Z_t is Markovian.

Second, we ask whether our CM-DTSM still matches the cross-section and the predictability evidence as we move from $\mathcal{N} = 5$ to $\mathcal{N} = 3$ sources of risk. The answer is yes. Reducing the number of shocks still leaves $2 \times \mathcal{N} = 6$ state variables—with 3 spanning variables to price the cross-section of yields and 3 forecasting variables to describe the dynamics of future yields—and produces little or no loss in terms of the variability and predictability of in excess returns. We consider reducing the dimension of the moving-average component, given the results based on Cochrane-Piazzesi predictive regressions. We find that LR test-statistics and BIC model selection criteria robustly support including a single factor driving the moving-average component. This single factor summarizes the information contained in the history of yields but that is not spanned by current yields. In practice, our three-factor CM-DTSM model produces yield decompositions that are economically different, with risk premium that are more cyclical. In other words, the expectation component plays a lesser role in explaining yield changes over the business cycles.

Third, we compare an unrestricted CM-DTSM with closely related results in Duffee (2011) based on a model with five Markovian factors, but where one "hidden" risk factor is filtered from the history of yields and generate substantial risk-premium variability. Non-Markovian risk factors and "hidden" risk factors may coexist in our framework and it is natural to check whether it is the case in the data. Consistent with Duffee (2011), filtering by itself adds to the predictability of the risk premium in 5-factor models. But including a moving-average component provides a better fit. In addition, we differ from Duffee (2011), since we show that it is unnecessary to increase the number of shocks (from three to five in his case). Instead, breaking from the assumption of Markovian risk factors introduces distinct spanning and forecasting roles for $\mathcal{P}_t^o$ and $\mathcal{E}_{\mathcal{P},t}$, with no additional shocks. Still, we agree with Duffee (2011) when he "advocates a significant change in the construction and estimation of multifactor term structure models" based on the information contained in the history of yields.

We contribute to several streams of the literature. First, we extend seminal results in Dai and Singleton (2002) showing that a 3-factor Gaussian DTSM matches the information content of current forward rates for future bond returns (Campbell and Shiller, 1991). Here, we find that a three-factor Gaussian CM-DTSM including a moving-average component captures the information in the history of forward rates.

Second, we add to the literature on DTSMs with unspanned risk factors, where macro variables or survey forecasts can help predict future returns but are not spanned by current yields (Chernov and Mueller, 2011; Joslin et al., 2014). The mechanism underlying unspanned risk is very different than ours. Additional surveys or macro variables are introduced by increasing the number of shocks. Then, the spanning restrictions in Joslin, Priebsch, and Singleton (2014) correspond to a projection of the economy-wide pricing kernel on "the priced risks in the bond market and on the state of the economy" (see p.1204). Once this restriction is imposed on the prices of risk, then the macro variables are unspanned by current yields. This mechanism also plays an important role to (nearly) hide one of the additional risk factors in Duffee (2011). In contrast, we introduce no additional sources of risk to generate separate spanning and forecasting roles. Hence, we do not project the pricing kernel on a subset of the shocks and we do not restrict the prices of risk. Importantly, a hidden factor is not precluded in our framework.

Third, we expand the applicability and relevance of Conditional Mean dynamics to model interest rates. Feunou and Fontaine (2014) introduce a macro-finance model where macro variables have Conditional Mean dynamics, where yields may span expected inflation but not inflation itself. This approach offers one resolution to the puzzle of unspanned macro variables. Here, we allow for Conditional Mean dynamics in yields to capture the predictive content in the history of forward rates. We discuss alternative strategies to integrate macro variables within this new framework, including adding unspanned macro variables.

Fourth, we expand the class of existing Gaussian DTSMs where risk factors have VAR(p) dynamics (with $p < \infty$).[3] Almost all empirical applications have $p = 1$ with two notable exceptions. Monfort and Pegoraro (2007) consider models with lags (as well as regimes) but limit their empirical section to cases with $p = 2$. Joslin, Le, and Singleton (2013) consider VAR(p) models with macro variables and estimate a model with $p = 3$.[4] But forecasting

[3] The usual argument is that longer lags can be recast within an extended state representation which has only one lag. See, e.g., Equations (7)-(8) in Ang and Piazzesi (2003) or footnote 7 in Joslin et al. (2011).

[4] Joslin, Le, and Singleton (2013) assume that risk factors have VAR(1) dynamic under $\mathbb{Q}$—as we do.

the bond risk premium requires longer lags of forward rates. We treat the case $p = \infty$ parsimoniously: our CM has the same number of parameters as a $VAR(2)$ (for $p > 2$, a VAR(p) has $(p-2)\mathcal{N}^2$ additional parameters). We find that out-of-sample forecast accuracy deteriorates quickly for $p > 1$. Even a short VAR(2) offers worse out-of-sample accuracy relative to a CM dynamics for yields.

Le, Singleton, and Dai (2010) discuss generalized prices of risk in the context of DTSMs. Our CM-DTSM is equivalent to generalizing the prices of risk to an affine function of Z_t and $\mathcal{E}_{Z,t-1}$. This generalization is consistent with equilibrium models where agents learn about the underlying process or where agents use 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 and we offer a risk-based reconcilation with the data. These results may be consistent with both compensation for risk and informational frictions.

Section 1 presents 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 the results.

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. We assess the contribution of current and past forward rates to predict bond returns for holding periods between one month and one year. We find (i) that a single combination of current forward rates summarizes their predictive content (ii) that long lags of the forward rates factor add to the predictability and (iii) that the weights on the current and past forecasting factors have decaying weights. Overall, the predictability evidence is consistent with adding a parsimonious moving-average component to standard VAR(1) dynamics for yields.

1.1 Cochrane-Piazzesi Regressions

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

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

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.[5] 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 Equation (1). The predictability increases 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 produce essentially the same results.

1.2 Cochrane-Piazzesi Regressions with Distributed Lags

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

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

different bond maturities and extend CP's regression to include distributed lags of the forward rate factor. Allowing for eight lags,

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

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.[6] 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 of interest rates (see, e.g., Koyck 1954; Griliches 1967). We adapt a simple distributed-lag model to the context of return predictability:

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

α_h is a maturity-specific scalar, with $0 \leq \alpha_h < 1$, and nesting Equation (1) if $\alpha_h = 0$. For any maturity and returns horizons, $\mathcal{R}_{t,h}$ is risk premium factor (with loadings $b_{n,h}$), with weight $1 - \alpha_h \gamma'_h$ on current forward rates, while α_h controls the weights on past forward rates. Substituting backward yields the distributed-lag representation:

$$\begin{aligned} xr^{(n)}_{t,h} &= \sum_{j=0}^{\infty} a_j \gamma'_h f_{t-j} + u^{(n)}_{t,h} \\ a_j &= b_{n,h} (1 - \alpha_h) \alpha_h^j = a_{j-1} \alpha_h. \end{aligned} \tag{4}$$

Equation (3) captures the decaying patterns observed in the estimates from Equation (2). Beyond its simplicity, this representation is also consistent with our term structure model (see Section 2.7).

Equation (3) can be estimated easily via non-linear least squares. We set the initial value $\mathcal{R}_{0,h}$ to the population mean. We also impose $N^{-1} \sum_{n=1}^{N} b_{n,h} = 1$ to identify $b_{n,h}$ from γ_h and α_h (N is the number of bond maturity in the cross-section of returns). Panel (B) of

[6] As in CP, we proceed via iterated OLS estimation over γ and a. Results are similar to their Table 5B.

Table 1 reports the R^2s from Equation (3). 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 6% and from 6% to around 11%, respectively, and increases substantially at longer horizons. Consistent with Duffee (2011), lagged forward rates are relatively more important for short-horizon returns.

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 $\alpha_{n,h}$ range 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 $b_{n,h}$. The estimates are centered around one by construction and increase with the bond duration. Panel (C) reports estimates of γ_h. These clearly exhibit the tent shape at every horizon.[7]

One natural question is whether we can collapse the return-forecasting factors $\gamma_h' f_t$ into an encompassing factor $\gamma' f_t$ that does not depend on the horizon. Indeed, we find that the horizon-specific factors are highly correlated. Table 3 reports results from a principal-component analysis of return-forecasting factors. The first principal component explains 97% of the total variation, with loadings that are spread out evenly across maturities. The contribution of the second component is small, but its loadings exhibit a slope patter across horizons.[8]

1.3 Yield Dynamics

The previous section uses predictability evidence to emphasize the importance of lagged forward rates. This section presents time-series evidence based on the joint dynamics of yields. 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. First, consider the following dynamics for Y_t at the monthly frequency:

$$\Delta Y_t = K_{Y0} + K_{Y1} Y_{t-1} + w_t, \qquad w_t = \epsilon_{Y,t} - \theta_Y \epsilon_{Y,t-1}, \tag{5}$$

[7] 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, as was the case for Equation (1).

[8] In the case $\alpha = 0$, the first principal component explains 93.7% of the total variations and the second component explains most of the remaining variations.

where $\epsilon_{Y,t}$ is a mean-zero innovation. Equation 5 is well-suited to match the predictability presented so far. The pattern of decaying coefficients in Equations (2) and (3) is consistent with a moving-average component in the yield innovation. In the context of time-series models, it is well known that inverting a moving-average term generates an autoregressive form with long lags and with decaying coefficients (see, e.g, Hamilton 1994). In the context of forecasting returns, Appendix A.1 shows that the pattern of decaying time-series coefficient translates in a pattern of decaying predictability coefficients. Second, it is natural to also consider VAR(p) specifications:

$$\Delta Y_t = K_{Y0} + K_{Y1} Y_{t-1} + \ldots + K_{Yp} Y_{t-p} + w_t^{(p)}. \tag{6}$$

For instance, Joslin et al. (2013) introduces MTSMs where bond yields and macroeconomic variables have VAR(p) dynamics. Also, CP use 12 lags to capture the predictability evidence from bond returns.[9]

1.3.1 Model Selection Criteria

We estimate Equation 5 as well as Equation 6 with $p = 1, \ldots, 12$. We report standard information criteria to assess which models provides the best *parsimonious* fit of the data (with 12 lags, as in CP, estimation involves 275 parameters!). Figure 1 shows the Hannan-Quinn (HQ) criteria and the BIC criteria in Panel A and B, respectively. The results unambiguously favor the inclusion of a moving-average component. The information criteria is lower for Equation 5 than for every VAR(p) model. This confirms the evidence above that the long-lag dependency observed in yields can be summarized parsimoniously using a moving-average component. Looking within VAR(p) models, the criteria favor a specification with $p = 3$ lags.[10]

[9] They use twelve lags for the purpose of constructing bootstrapped inference

[10] The HQ criteria reaches a minimum at $p = 3$ while the BIC criteria favors 2 or 3 lags (the minimum is achieved for $p = 2$ but the numerical values are very close). The AIC criteria is well-known to favor over-parameterized specifications in finite samples. We do not report the AIC criteria, since this bias appears severe in our sample: the AIC criteria reaches a minimum for $p = 14$. See Lütkephol 2006, 4.3 for a discussion of model selection and a comparison of information criteria in the multivariate case.

1.3.2 Residuals Diagnostics

There is also a direct way to reveal a moving-average component hidden in the estimated residuals of a VAR(1). If that is the case, then $w_t^{(1)}$ is correlated with $w_{t-1}^{(1)}$, since $\theta_Y \neq 0$, but not with longer lags. We can easily estimate these correlations based on the estimated residuals $\hat{w}_t^{(1)}$. First, we focus on the auto-correlation for each component of $w_t^{(1)}$. Panel (A) of Table 4 reports the results from regressions of each element $w_{n,t}^{(1)}$ on several of its own lags $w_{n,t-j}^{(1)}$. We find that the first lag is significant but that longer lags are insignificant, consistent with a moving-average component with one lag. Next, Panel (B) reports the cross-correlations between each element of $w_t^{(1)}$ and each element of its lag $w_{t-1}^{(1)}$. Again, the correlation coefficients are significant at the 5% level in the majority of cases, and significant at the 10% level in almost all cases. Repeating these diagnostics using the residuals $\epsilon_{Y,t}$ from Equation 5 produces insignificant coefficients in all cases (unreported).

We also perform a formal statistical test of the null hypothesis that the estimated residuals are white noise. We apply the Breusch-Godfrey Lagrange Multiplier test for serial correlation to every model. In the case of a VAR(1), this test is based on the following auxiliary regression:

$$\hat{w}_t^{(1)} = a_1 Y_{t-1} + a_2 \hat{w}_{t-1}^{(1)} + \eta_t, \tag{7}$$

where all coefficients should be zero if $\hat{w}_t^{(1)}$ is a white noise. We use the original LM statistics and the modified LMF statistics (Godfrey, 1978; Doornik, 1996).[11] The p-values are essentially zero in both cases for the null hypothesis of no autocorrelation. The null hypothesis is not rejected for other models (unreported).

1.3.3 Factor Structure

In addition, the estimated residuals reveal a factor structure: each component of w_t is correlated with essentially the same linear combination of w_{t-1}. We adopt the following two-step estimation procedure to extract the common factor. First, we regress the cross-

[11] For VAR(p) models, the auxiliary regression also includes any additional lags Y_{t-i}. The original test-statistics have an asymptotic χ^2 distribution while the alternative statistics in Doornik (1996) have an F distribution. Both test-statistics yield similar results in our sample. Note that the well-known portmanteau test for serial correlation is inappropriate when the regression includes lagged depended variables. E.g., see Hayashi (2001) (Section 2.10) and Maddala (2001).

section average $\bar{w}_t \equiv N^{-1} \sum w_t^{(n)}$ on w_{t-1}:

$$\bar{w}_t = \gamma_\epsilon' \bar{w}_{t-1} + \eta_t,$$

and then regress each element $w_t^{(n)}$ on $\hat{\gamma}_w' w_{t-1}$. These two steps provide an estimate of the single-factor regression:

$$w_t = \alpha \gamma_w' \bar{w}_{t-1} + \eta_t, \tag{8}$$

with the identification restriction that $\sum_n \alpha_n = N$.[12] Imposing the factor structure, if true, should improve the precision of the estimates. Panels (C) and (D) of Table 4 report the results from the first and the second step, respectively. The estimate of γ_w shows the familiar tent shape. We can reject the null hypothesis that the vector of coefficients is jointly equal to zero: the F-statistic for the restriction $\gamma_w = 0$ has a p-value of 0.02. Finally, the coefficient estimates α_n on the lagged factor $\gamma_w' w_{t-1}$ are all statistically significant at the 1% level.

1.3.4 Out-of-Sample Comparison

The combination of information criteria, diagnostics based on residuals and specification strongly supports the parsimonious specification in Equation 5. Comparisons based on out-of-sample forecast performance only add to the evidence. We perform the following exercise. We first estimate Equation 5 as well as Equation 6 with $p = 1, 2, 3$ and $p = 12$ with data until the end of 1985. Standing at the beginning of 1986, we use these estimates to forecast Y_t for every month in 1986, up to twelve month ahead. Then, we re-estimate the model with data until the end of 1986 and we repeat the forecasting exercise for 1987. We repeat for every year until the end of our sample. Table 5 reports the Root Mean Squared Errors (RMSE) associated with each forecast horizons, expressed as a multiple of the RMSE for the benchmark VAR(1) model. The specification in Equation 5 leads to little changes in out-of-sample forecast accuracy. This is not true for other models. In particular, the VAR(2) model adds as many parameters as a moving-average, but the out-of-sample forecasts are unambiguously worse. Adding further lags only leads to further deteriorations.

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

2 Unspanned Information in the History of Yields

The previous section revisited and added to the evidence showing that lagged forward rates predict bond excess returns. This section develops a parsimonious extension of standard DTSMs where the current cross-section of yields is not sufficient to forecast future yields. We follow Duffee (2011) in the empirical implementation in allowing for measurement errors in all yields. We differ from Duffee (2011) and Joslin, Priebsch, and Singleton (2014) in that we do not rely on any price of risk restriction.

2.1 Risk-Neutral Dynamics

We postulate a small number of $\mathcal{N}$ 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 $Z_t \in \mathbb{R}^{\mathcal{N}}$,*

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

where the distribution of Z_t conditional on the history $\mathcal{I}_{Z,t-1} = \{Z_{t-1}, \ldots, Z_0\}$ is Gaussian with constant variance under the historical distribution $\mathbb{P}$.

Assumption 1 is consistent with the observation that a small number of principal components suffice to explain the variations in the term structure of yields. Note that that the coefficients A_n and B_n are unrestricted at this stage. For completeness, and to be clear about what we mean that some random variable is Markovian, we provide the following standard definition of the Markov property.

Definition 1 *We say that Z_{t+h} is Markovian (has the Markov property) under some measure $\mathbb{P}$ if its conditional distribution given its own history $\mathcal{I}_{Z,t} = \{Z_t, \ldots, Z_0\}$ only depends on its current value:*

$$\mathbb{P}(Z_{t+h} = z | \mathcal{I}_{Z,t}) = \mathbb{P}(Z_{t+h} = z | Z_t) \qquad h > 0. \tag{10}$$

Note that Assumption 1 does not imply that Z_t is Markovian under $\mathbb{P}$ since (i) its conditional mean can depend on the whole history and (ii) its conditional distribution for $h > 1$ is left unspecified. Proposition 1 clarifies an important implication of the absence of arbitrage under Assumption 1: the risk factors Z_t must be Markovian under the risk-neutral measure $\mathbb{Q}$.

Proposition 1 *If bond prices offer no arbitrage opportunity and if the conditional distribution of $Z_t \in \mathbb{R}^{\mathcal{N}}$ under the risk-neutral distribution $\mathbb{Q}$ is Gaussian conditional on $\mathcal{I}_{Z,t-1} = \{Z_{t-1}, \ldots, Z_0\}$ and with constant variance, then Assumption 1 is equivalent the Markov property for Z_t under $\mathbb{Q}$ and therefore, we have that:*

$$\Delta Z_t = K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}} Z_{t-1} + \Sigma \epsilon_t^{\mathbb{Q}}, \tag{11}$$

where $\epsilon_t^{\mathbb{Q}}$ is a standard Gaussian innovation with respect to $\mathcal{I}_{Z,t-1}$. In addition, A_n and B'_n follow (standard) recursions given in Appendix A.2 with initial values $A_1 = \rho_0$ and $B_1 = \rho_1$.

One important implication is that, for the class of models we consider, the information content in the history of yields is not an implication of no-arbitrage. This result should not come as a surprise. After all, nearly all GDTSMs use dynamics as in Equation (11) to derive Equation (9) for yields. This is also consistent with the model-free evidence in Joslin et al. (2013) that pricing factors are Markovian under the $\mathbb{Q}$ measure. Nonetheless, Proposition 1 is useful in clarifying that the implication works the other way as well. Once we fix the number of linear risk factors required to explain the cross-section of yields, then these factors must have linear Markovian dynamics under any risk-neutral measure. The result is not trivial but depends crucially on the absence of arbitrage opportunities among bond prices.

2.2 Historical Dynamics

If Z_t is Markovian under $\mathbb{Q}$, then how can the no-arbitrage restriction be consistent with the stylized fact that *past* yields predict excess returns? To see that it is, consider the excess returns from holding an n-period bond between t and $t+h$,

$$xr_{t,h}^{(n)} = -(n-h)y_{t+h}^{(n-h)} + ny_t^{(n)} - hy_t^{(h)}.$$

Under the conditions of Assumption 1, this is given by

$$xr_{t,h}^{(n)} = \mathcal{B}'_{n-h} Z_{t+h} + (\mathcal{B}_h - \mathcal{B}_n)' Z_t + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n, \tag{12}$$

where $\mathcal{A}_n = -nA_n$ and $\mathcal{B}_{n-1} = -nB_n$. Taking conditional expectations under the $\mathbb{P}$-measure, the bond risk premium $brp_{t,h}^{(n)} \equiv E^{\mathbb{P}}[xr_{t,h}^{(n)} | \mathcal{I}_{Z,t}]$ is given by

$$brp_{t,h}^{(n)} = \mathcal{B}'_{n-h} E[Z_{t+h} | \mathcal{I}_{Z,t}] + (\mathcal{B}_h - \mathcal{B}_n)' Z_t + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n. \tag{13}$$

Therefore, Assumption 1 implies that the predictability of bond returns returns depends on multiple lags only if Z_{t+h} is not Markovian under the $\mathbb{P}$-measure. Assumption 2 introduces the CM dynamics under $\mathbb{P}$ where Z_{t+j} is not Markovian from the standpoint of investors factors directly.[13]

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

$$\begin{aligned} Z_t &= \mathcal{E}_{Z,t-1}^{\mathbb{P}} + \Sigma \epsilon_t^{\mathbb{P}} \\ \Delta \mathcal{E}_{Z,t}^{\mathbb{P}} &= K_0^{\mathbb{P}} + K_1^{\mathbb{P}} \mathcal{E}_{Z,t-1}^{\mathbb{P}} + \Sigma_{\mathcal{E}_Z} \epsilon_t^{\mathbb{P}}, \end{aligned} \tag{14}$$

where $\mathcal{E}_{Z,t}^{\mathbb{P}} \equiv E^{\mathbb{P}}[Z_{t+1} | \mathcal{I}_{Z,t}]$ and $\epsilon_t^{\mathbb{P}}$ is a standard Gaussian $(\mathcal{N} \times 1)$ vector of innovations relative to $\mathcal{I}_{Z,t-1}$. To see the link between Equation (11) *and Equation* (14)*, regroup and*

[13]If we observe Z_t with errors, the expectation of Z_{t+h} conditional on the econometrician's information set will depends on the past via the filtering step irrespective of the true dynamics. We include this additional mechanism in the Section 2.3.

substitute for $\mathcal{E}^{\mathbb{P}}_{Z,t-1}$ *on the right-hand side of Equation* (14)*:*

$$\begin{aligned}
\mathcal{E}^{\mathbb{P}}_{Z,t} &= K^{\mathbb{P}}_0 + (K^{\mathbb{P}}_1 + I_{\mathcal{N}})(Z_t - \Sigma\epsilon^{\mathbb{P}}_t) + \Sigma_{\mathcal{E}_Z}\epsilon^{\mathbb{P}}_t \\
&= K^{\mathbb{P}}_0 + (K^{\mathbb{P}}_1 + I_{\mathcal{N}})Z_t - \Big((K^{\mathbb{P}}_1 + I_{\mathcal{N}})\Sigma - \Sigma_{\mathcal{E}_Z}\Big)\,\epsilon^{\mathbb{P}}_t \\
&= K^{\mathbb{P}}_0 + (K^{\mathbb{P}}_1 + I_{\mathcal{N}})Z_t - \Big((K^{\mathbb{P}}_1 + I_{\mathcal{N}}) - \Sigma_{\mathcal{E}_Z}\Sigma^{-1}\Big)\,\Sigma\epsilon^{\mathbb{P}}_t,
\end{aligned}$$

so that the dynamics of Z_t *under* $\mathbb{P}$ *is given by*

$$\Delta Z_t = K^{\mathbb{P}}_0 + K^{\mathbb{P}}_1 Z_{t-1} - \theta\Sigma\epsilon^{\mathbb{P}}_{t-1} + \Sigma\epsilon^{\mathbb{P}}_t,$$

where we defined

$$\theta \equiv K^{\mathbb{P}}_1 + I_{\mathcal{N}} - \Sigma_{\mathcal{E}_Z}\Sigma^{-1}. \tag{15}$$

and Z_t *is Markovian (as in Equation* (11)*) if and only if* $\theta = 0$*.*

We discuss Assumption 2 in details.

Spanning and Forecasting

The states Z_t and $\mathcal{E}_{Z,t}$ play distinct roles. The risk 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 mean $\mathcal{E}_{Z,t}$ is a function of the entire history of Z_t (iterate Equation (14) backward) and summarizes the evolution of future yields. This separation of the *spanning* and *forecasting* roles is at the heart of our modelling strategy.

This separation arises because the states Z_t and $\mathcal{E}^{\mathbb{P}}_{Z,t}$ are driven by the *same* $\mathcal{N}$ shocks $\epsilon^{\mathbb{P}}_t$. This is uncontroversial: $\mathcal{E}^{\mathbb{P}}_{Z,t}$ is a deterministic function of Z_t in a VAR(1). In the case of CM-DTSM, the Z_t innovations are given by $\Sigma\epsilon^{\mathbb{P}}_t$, while the $\mathcal{E}^{\mathbb{P}}_{Z,t}$ innovations are given by $\Sigma_{\mathcal{E}_Z}\epsilon^{\mathbb{P}}_t$. While the number of state variables increases from $\mathcal{N}$ to $2\mathcal{N}$, the covariance matrix of Z_t and $\mathcal{E}^{\mathbb{P}}_{Z,t}$, jointly, has rank $\mathcal{N}$ only (conditionally or unconditionally as estimated via principal-component analysis, say).

Macro-Finance Term Structure Models

Feunou and Fontaine (2014) introduce the family of macro-finance term structure models

(CM-MTSMs) where macro variables have CM dynamics and yields span the conditional mean of macro variables (but not the macro variables themselves). On the other hand, yields have standard VAR(1) dynamics. In fact,there is no difference between a standard DTSM and a CM-DTSM once observations of macro variables are removed from the specification. Without macro variables, yield portfolios follow a VAR(1) under $\mathbb{Q}$ and $\mathbb{P}$, exactly as in Joslin, Singleton, and Zhu (2011). In other words, Feunou and Fontaine (2014) is a "macro-finance" paper while our case is a "yield-only" paper.

To introduce macro variables in our framework, one natural approach is to follow Joslin, Priebsch, and Singleton (2014) and suppose that macro variables are linear function of the latent factor Z_t. One may then impose the restrictions of unspanned macro variables, as in Joslin, Priebsch, and Singleton (2014), or let the data speak, as in Chernov and Mueller (2011).[14] In either case, this approach involves the following trade-off relative to Feunou and Fontaine (2014). Forecasts of the macro variables may be improved, since they are based on a rich information set containing the history of yields and macro variables. However, pricing errors for yields may increase since spanning restrictions reduce the dimension of the risk factors generating yields. Adding risk factors may offer one resolution of this trade-off.

Markovian Expectations

Assumption 2 is equivalent to an unrestricted VARMA(1,1) for Z_t where θ is the moving-average parameter. This is consistent with the evidence in Section 1. However, the representation in Equation (14) is more transparent. It shows that the conditional mean $\mathcal{E}^{\mathbb{P}}_{Z,t}$ is Markovian under $\mathbb{P}$: $\mathcal{E}^{\mathbb{P}}_{Z,t}$ depends on lagged conditional mean $\mathcal{E}^{\mathbb{P}}_{Z,t-1}$ via the auto-regressive matrix $(I_{\mathcal{N}} + K^{\mathbb{P}}_1)$ and on the innovation $Z_t - \mathcal{E}^{\mathbb{P}}_{Z,t-1}$ via the matrix $\Sigma_{\mathcal{E}_Z}\Sigma^{-1}$.[15] This is analogous to a GARCH(1,1) model where the conditional variance can be filtered directly from observed (squared) returns innovations. The risk factors Z_t are Markovian when these weights are equal. In addition, the conditional mean $\mathcal{E}^{\mathbb{P}}_{Z,t}$ can be filtered easily given the

[14] See Appendix B in Joslin, Priebsch, and Singleton 2014. The restriction ties the parameters governing the prices of the risks ϵ_t with the parameters governing the dynamics of Z_t under the historical measure $\mathbb{P}$.

[15] In principle, we could generalize Equation 14 to include additional lags $\mathcal{E}^{\mathbb{P}}_{Z,t-j}$ and $Z_{t-j} - \mathcal{E}^{\mathbb{P}}_{Z,t-j-1}$. This would generate VARMA(p,q) processes for Z_t. We neglect this generalization because (i) the evidence in Section 1 suggests that one lag is well-specified, (ii) general (p,q) specifications are notoriously complex to check for identification and (iii), for reasons that will become clear in Section 2.5, these identification issues are still unresolved in cases where Z_t is latent.

time-series of $Z_t^{\mathbb{P}}$ using the recursion in Equation (14) and starting from some initial value $\mathcal{E}_{Z,0}^{\mathbb{P}}$. Nonetheless, Z_t is typically latent and must be filtered from the observed yields, a point which we address in Section 2.3 below. Finally, note that the conditional expectation is not Markovian when Z_t follows a VAR(p) with $p < \infty$.

General Equilibrium

Expectations and current values are driven by the same set of shocks in DSGE models; 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

The statistical properties of observed yields are required to complete the model specification. 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. We consider the case where measurement errors are pervasive.

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

One can replace $\mathcal{P}_t$ with Y_t through the rest of the paper if $W = I$, the identity matrix. Alternatively, W can be the mapping from yields to forward rates (as we do below for estimation) or it can be derived from principal component analysis. Assumption 3 leads to a well-known Kalman filtering problem, which is often summarized via the following state-space representation:

$$\mathcal{P}_t^o = H_0 + H' Z_t + w_t \tag{16}$$

$$Z_t = \mathcal{E}_{Z,t-1}^{\mathbb{P}} + \Sigma \epsilon_t^{\mathbb{P}}. \tag{17}$$

Equation (16) stacks the measurement equations for yields and Equation (17) corresponds to the state dynamics. This representation is based on the latent factors Z_t, which need to be filtered. Time-series and cross-section implications from the model are not directly 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 estimates of Z_t and $\mathcal{E}_{Z,t}^{\mathbb{P}}$ filtered from the observed history $\mathcal{I}_{\mathcal{P},t} = \{\mathcal{P}_t^o, \mathcal{P}_{t-1}^o, \ldots\}$.

Theorem 1 establishes that state-space model given by Equations (16)-(17) belong to the family of CM-DTSMs and can be represented directly in terms of the observed $\mathcal{P}_t^o$ and its conditional mean $\mathcal{E}_{\mathcal{P},t}$. The key result in Theorem 1 is the mapping between the new parameters and the parameters for the latent dynamics. Hence, the model provides all the implications from the observable portfolios relevant for the underlying risk factors.[16]

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 $R \equiv var(\mathcal{P}_t - \mathcal{P}_t^o)$, a full-rank diagonal matrix, and $\mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} \equiv E^{\mathbb{P}}[\mathcal{P}_{t+1}^o | \mathcal{I}_{\mathcal{P},t}]$. Then, $\mathcal{P}_t^o$ has the following conditional mean dynamics:*

$$\mathcal{P}_t^o = \mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}} + \Sigma_{\mathcal{P}} \epsilon_{\mathcal{P},t}^{\mathbb{P}} \tag{18}$$

$$\Delta \mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} = K_{0\mathcal{P}}^{\mathbb{P}} + K_{1\mathcal{P}}^{\mathbb{P}} \mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}} + \Sigma_{\mathcal{E}_{\mathcal{P}}} \epsilon_{\mathcal{P},t}^{\mathbb{P}}. \tag{19}$$

[16] Theorem 1 builds on a fundamental result from Lütkephol (1984): a linear transformation of VARMA processes also follows a VARMA process. However, Lütkephol (1984) does not characterize the parameters of this new process.

Define $H^{\perp}H' = I_{\mathcal{N}}$, then the parameters in Equations (18)-(19) *are given by*

$$\begin{aligned} K^{\mathbb{P}}_{1\mathcal{P}} &= H'(K^{\mathbb{P}}_{1} + I_{\mathcal{N}})H^{\perp} - I_{\mathcal{J}} \\ K^{\mathbb{P}}_{0\mathcal{P}} &= -K^{\mathbb{P}}_{1\mathcal{P}}H_0 + H'K^{\mathbb{P}}_{0} \\ \Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}} &= P + R \\ \Sigma_{\mathcal{E}_{\mathcal{P}}} &= \Big((K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}})P - H'\theta\Sigma\Sigma'H\Big)(\Sigma'_{\mathcal{P}})^{-1}, \end{aligned} \tag{20}$$

where the eigenvalues of $K^{\mathbb{P}}_{1}$ are all inside $(-2,0)$ and P is a $\mathcal{J} \times \mathcal{J}$ matrix with rank $\mathcal{N}$ and given by the solution to a generalized algebraic Ricatti equation:

$$P = A'PA - (A'P + S)('P + R)^{-1}(PA + S') + Q$$

with $A = (K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}})'$ and

$$\begin{aligned} Q &= -A'\left(H'\Sigma\Sigma'H\right)A + H'\Sigma\Sigma'H + H'\Sigma_{\mathcal{E}}\Sigma'_{\mathcal{E}}H \\ S &= -A'H'\Sigma\Sigma'H \ + H'\Sigma_{\mathcal{E}}\Sigma'H. \end{aligned}$$

Theorem 1 is similar in spirit to Theorem 1 in Joslin, Singleton, and Zhu (2011).[17] Importantly, Theorem 1 summarizes all that we can really know about the yield dynamics given the observable data $\mathcal{I}_{\mathcal{P},t}$. We recover the true parameter of the portfolio dynamics conditional on observing $\mathcal{I}_{\mathcal{P},t}$. The theorem also delivers estimates of the latent factor Z_t based on the observable history $\mathcal{I}_t$. The proof uses the steady-state solution of the Kalman filter to connect the state-space representation in Equations (16)-(17) to the direct representation in Equations (18)-(19).[18] This representation is in the spirit of the innovation representation of equilibrium models in Hansen and Sargent (2014).

[17] 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 here: their canonical form requires an unrestricted VAR specification.

[18] Implementation of the state-space representation in Equations (16)-(17) for standard term structure models shows that the filter converges very fast to its steady state in practice.

2.4 Two Mechanisms for Unspanned Information

Theorem 1 identifies two mechanisms introducing a role for the history of yields in the dynamics of the observed portfolios $\mathcal{P}_t^o$. This is one clear benefit from the representation in Theorem 1. Apply the definition in Equation 15 to the $CM-DTSM$ in Theorem 1 and subsitute using the parameters in Equation 20:

$$
\begin{aligned}
\theta_{\mathcal{P}} \equiv\; & K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}} - \Sigma_{\mathcal{E}_{\mathcal{P}}}\Sigma_{\mathcal{P}}^{-1} \\
=& K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}} - \Big((K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}})P - H'\theta\Sigma\Sigma' H\Big)\left(\Sigma'_{\mathcal{P}}\right)^{-1}\left(\Sigma_{\mathcal{P}}\right)^{-1} \\
=& K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}} - \Big((K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}})\left(\Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}} - R\right) - H'\theta\Sigma\Sigma' H\Big)\left(\Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}}\right)^{-1} \\
=& (K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}})R\left(\Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}}\right)^{-1} + H'\theta\Sigma\Sigma' H\left(\Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}}\right)^{-1} \qquad (21)
\end{aligned}
$$

The portofolios $\mathcal{P}_t^o$ are not Markovian if and only if either the latent risk factor Z_t is non-Markovian—the first term—or the portfolios are measured with errors—the second term—or both. The first term in Equation (21) is zero if $R = 0$. This term is consistent with CP's suggestion 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" and that, using the Kalman filter, "the best guess of the true [state] is a geometrically weighted moving average." (See Cochrane and Piazzesi 2005, p. 154.). This term is also the source of the hidden factor in Duffee (2011): the conditional dynamics of $\mathcal{P}_t^o$ depend on the history of yields via the Kalman filter. Empirically, the dynamics of $\mathcal{P}_t^o$ 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 $\Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}}$.[19] This mechanism is restricted because R is a diagonal covariance matrix ($\mathcal{N}$ parameters) to preserve the interpretation of the measurement errors. Furthermore, the more restricted case $R = \sigma^2 I$ is often used in practice (Duffee, 2011).

The second term in Equation (21) provides a more flexible channel to introduce a role for the history of yields. Since measurement errors are small, the innovations in $\mathcal{P}_t$ represent a large share of the innovations in $\mathcal{P}_t^o$: there is no a priori reason that $H'\theta\Sigma\Sigma' H$ should be

[19] The dynamics would close to Markovian if the persistence of the yield portfolios $(K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}})$ is small, but this case is irrelevant given the persistence observed in the data.

small relative to $\Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}}$. This is in contrast with the measurement error mechanism where a relatively small value for R reduces the the role of past information. In addition, note that the rank of $\theta_{\mathcal{P}}$ is at most $\mathcal{N}$, the number of risk factors, irrespective of the number $\mathcal{J}$ of yield portfolios used at estimation.

2.5 Identification

The parameters to estimate (see Equation 20) are not jointly identified. Assumption 4 summarizes standard identification restrictions.

Assumption 4 *We impose the following identification restrictions:*

i) $K_0^{\mathbb{P}} = 0$ *and* $\Sigma = I_{\mathcal{N}}$*, to fix the level and the scale of the latent factors;*

ii) *the eigenvalues of* θ *remain within the unit circle, so that the process for* Z_t *is invertible;*

iii) *the polynomial matrices* $\mathcal{I}_{\mathcal{N}} - (K_1^{\mathbb{P}} + \mathcal{I}_{\mathcal{N}})z$ *and* $\mathcal{I}_{\mathcal{N}} - \theta_1^{\mathbb{P}} z$ *have no common roots.*

Therefore, the parameters of interest are $\Theta^{\mathcal{P}} = \{H_0, H, K_1^{\mathbb{P}}, R, \theta\}$. These parameters a not separately identified yet. The interaction between a moving-average with measurement errors introduces an additional identification challenge. Consider the simplest case where the state variable follows a univariate MA(1) process and where the measurement equation is also scalar. Then, existing results (Hamilton, 1994) show that only 2 parameters between the loading, variance and moving-average (scalar) can be separately identified; H, R, and θ respectively in our case. But more can be identified in the general case.

In the general case, the key equations to study the tradeoffs are

$$
\begin{aligned}
H'\theta\theta' H + H'H &= \left(\theta_{\mathcal{P}}\Omega_{\mathcal{P}}\theta'_{\mathcal{P}} + \Omega_{\mathcal{P}}\right) - \left(\Phi_{1\mathcal{P}}^{\mathbb{P}} R \Phi_{1\mathcal{P}}^{\mathbb{P}\prime} + R\right) \\
H'\theta H &= \theta_{\mathcal{P}}\Omega_{\mathcal{P}} - \Phi_{1\mathcal{P}}^{\mathbb{P}} R.
\end{aligned}
$$

Jointly, these equations characterize exactly what we can learn from the data about H, θ and R based on the reduced-form estimates $\Phi_{1\mathcal{P}}^{\mathbb{P}}$, $\Omega_{\mathcal{P}}$ and $\theta_{\mathcal{P}}$. The cases $R = 0$ and $\theta = 0$ correspond to known results. If $R = 0$, there is no measurement error and we can show that θ and $H'H$ are identified. If $\theta = 0$, there is no moving-average term and we can identify a diagonal variance matrix R and $H'H$. But Proposition 2 shows that these extreme cases are overly restrictive.

Proposition 2 *Given Assumption 4, and suppose*

i) *θ is symmetric and*

ii) *R is diagonal with positive elements,*

then H_0, $K_1^{\mathbb{P}}$, R, θ and $H'H$ are identified. Rewrite $H = [H_1 \; H_2]$ where H_1 is a $\mathcal{N} \times \mathcal{N}$ square matrix. Suppose H_1 is upper triangular with positive diagonal elements, then H_1 and H_2 are identified.
See the proof in the online appendix.[20]

From Proposition 2, the CM-DTSM parameters $\Theta^{\mathcal{P}} = \{H_0, H, K_1^{\mathbb{P}}, R, \theta\}$ are identified. This is a novel result in itself, opening the way to identify separate roles for measurement errors and for a moving-average component in the time-series of yields. Restricting R to a diagonal variance matrix is a natural and standard choice. Restricting θ to a symmetric matrix is easy to implement in practice and flexible enough to accommodate the data.[21] In fact, the evidence in Section 1.3.3 suggests that $\theta_{\mathcal{P}}$ has rank 1 or 2. We explore this possibility in Section 3, showing that further restrictions on the rank of θ are supported in the data.

2.6 Risk-Neutral Parameters

The parametrization $\Theta^{\mathcal{P}} = \{H_0, H, K_1^{\mathbb{P}}, R, \theta\}$ does not include parameters of the $\mathbb{Q}$-dynamics for Z_t. Instead, we allow for free parameters in H to emphasize that our theoretical results do not rely on no-arbitrage. Nonetheless, the generic factor model $\mathcal{P}_t = H_0 + H'Z_t$ in Equation (16) imposes cross-equation restrictions due to the role of the loading matrix H in the parametrization given by Equation (20).

[20]The online appendix is available on the authors' home pages: `http://kamkui.net/docs/ConditionalMeanTermStructureModels_OnlineAppendix.pdf`.

[21]Consider the case $\mathcal{N} = 3$. Is easy to generate a symmetric θ starting from any lower triangular $\mathcal{N} \times \mathcal{N}$ matrix T: $\theta = TT'$. In this case, there are 6 free parameters to estimate. To implement the restriction that θ has rank 2, drop the last column of T. In this case there are 5 parameters to estimate. For the case with a rank of 1, drop the last two column of T, keeping only 3 parameters to estimate in the first column.

CM-DTSMs can be estimated as a factor model for yields without imposing no-arbitrage. This would let the data impose no-arbitrage since observed bond prices are free of arbitrage (Diebold and Li, 2006). However, the vector H_0 and the matrix H embody additional restrictions if bond prices are constructed from a no-arbitrage argument:

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

where A_n and B'_n are given in terms of the parameter $K_0^{\mathbb{Q}}$, $K_1^{\mathbb{Q}}$, ρ_0 and ρ_1 (see Proposition 1). We follow Dai and Singleton (2000) to identify parameters of the risk factor dynamics under $\mathbb{Q}$. Specifically, ρ_0 and $K_0^{\mathbb{Q}}$ are unrestricted, the elements of ρ_1 are positive, and $K_1^{\mathbb{Q}}$ is lower triangular.[22] Summing up, adding no-arbitrage restrictions generates a different CM-DTSM specification with parameters given by

$$\widetilde{\Theta}^{\mathcal{P}} = \{\rho_0, \rho_1, K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, K_1^{\mathbb{P}}, R, \theta\}. \tag{23}$$

The risk-neutral parameters $\widetilde{\Theta}^{\mathcal{P}}$ are just identified if $\mathcal{J} = \mathcal{N} + 1$. Otherwise, if $\mathcal{J} > \mathcal{N} + 1$, the no-arbitrage conditions provide over-identifying restrictions.

2.7 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 3 derives the pricing kernel that is consistent with the $\mathbb{Q}$-dynamics and the $\mathbb{P}$-dynamics in Equations (11) and (14), respectively.

Proposition 3 *The unique pricing kernel M_{t+1} consistent with the $\mathbb{Q}$-dynamics and the $\mathbb{P}$-dynamics in Equations* (11) *and* (14) *is given by*

$$M_{t+1} \equiv \exp\left(-\frac{\Lambda_t' \Sigma \Sigma' \Lambda_t}{2} - \Lambda_t' \Sigma \epsilon_{t+1}^{\mathbb{P}}\right), \tag{24}$$

[22] As in Assumption 4, Dai and Singleton (2000) also take Σ as the identity matrix.

with prices of risk

$$\Lambda_t \equiv (\Sigma\Sigma')^{-1}\left(\Lambda_0 + \Lambda_Z Z_t + \Lambda_{\mathcal{E}} \mathcal{E}^{\mathbb{P}}_{Z,t-1}\right). \tag{25}$$

The mapping between parameters of Equations (11) *and* (14) *is given by*

$$\begin{aligned} K_0^{\mathbb{P}} &= \Lambda_0 + K_0^{Q} \\ K_1^{\mathbb{P}} &= K_1^{\mathbb{Q}} + \Lambda_Z + \Lambda_{\mathcal{E}} \\ \Sigma_{\mathcal{E}_Z} &= \left(K_1^{\mathbb{Q}} + I_{\mathcal{N}} + \Lambda_Z\right)\Sigma. \end{aligned} \tag{26}$$

Therefore, the CM dynamics for Z_t in Assumption 2 corresponds to a generalization of the prices of risk. Indeed, we have that Z_t is not Markovian if $\Lambda_{\mathcal{E}} = 0$, yielding the standard pricing kernel. Otherwise, the prices of risk are functions of both Z_t and $\mathcal{E}_{Z,t-1}$. See Le, Singleton, and Dai (2010) for similar generalizations. 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 4 confirms that long lags of Z_t enter the bond risk premium.

Proposition 4 *The bond risk premium from holding an n-period bond between t and $t+h$,*

$$brp_{t,h}^{(n)} \equiv E[xr_{t,h}^{(n)} | \mathcal{I}_{Z,t}], \tag{27}$$

is given by

$$brp_{t,h}^{(n)} = \delta_{h,0} + \delta_{h,Z}^{(n)\prime} Z_t + \delta_{h,\mathcal{E}}^{(n)\prime} \mathcal{E}^{\mathbb{P}}_{Z,t-1}, \tag{28}$$

where $xr_{t,h}^{(n)}$ is defined in Equation (12) *and with coefficients given by*

$$\begin{aligned} \delta_{h,0}^{(n)} &= \mathcal{B}_{n-h}'\left[(K_1^{\mathbb{P}} + I_{\mathcal{N}})^h - I_{\mathcal{N}}\right](K_1^{\mathbb{P}})^{-1} K_0^{\mathbb{P}} + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n, \\ \delta_{h,Z}^{(n)\prime} &= \mathcal{B}_{n-h}'(K_1^{\mathbb{P}} + I_{\mathcal{N}})^{h-1}\left(K_1^{\mathbb{P}} + I_{\mathcal{N}} - \theta\right) + (\mathcal{B}_h - \mathcal{B}_n)', \\ \delta_{h,\mathcal{E}}^{(n)\prime} &= \mathcal{B}_{n-h}'(K_1^{\mathbb{P}} + I_{\mathcal{N}})^{h-1}\theta. \end{aligned}$$

Conditioning on $\mathcal{I}_{\mathcal{P},t}$, the information generated by $\mathcal{P}_t^o$, the bond risk premium $\widehat{brp}_{t,h}^{(n)} =$

$E[brp_{t,h}^{(n)}|\mathcal{I}_{\mathcal{P},t}]$ *is given by*

$$\widehat{brp}_{t,h}^{(n)} \equiv \delta_{h,0\mathcal{P}}^{(n)} + \delta_{h,\mathcal{P}}^{(n)\prime}\mathcal{P}_t^o + \delta_{h,\mathcal{E}_\mathcal{P}}^{(n)\prime}\mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}}, \tag{29}$$

with

$$\begin{aligned}
\delta_{h,0\mathcal{P}}^{(n)} &= \mathcal{B}_{n-h}' \left((K_1^{\mathbb{P}} + I_\mathcal{N})^h - I_\mathcal{N}\right)(K_1^{\mathbb{P}})^{-1}K_0^{\mathbb{P}} + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n \\
&\quad - \left(\mathcal{B}_{n-h}'(K_1^{\mathbb{P}} + I_\mathcal{N})^{h-1}K_1^{\mathbb{P}} + (\mathcal{B}_h - \mathcal{B}_n)'\right) H^\perp H_0 \\
\delta_{h,\mathcal{P}}^{(n)\prime} &= \mathcal{B}_{n-h}'(K_1^{\mathbb{P}} + I_\mathcal{N})^{h-1}H^\perp \left(K_{1\mathcal{P}}^{\mathbb{P}} + I_\mathcal{J} - \theta_\mathcal{P}\right) + (\mathcal{B}_h - \mathcal{B}_n)' H^\perp P(P+R)^{-1} \\
\delta_{h,\mathcal{E}_\mathcal{P}}^{(n)\prime} &= \mathcal{B}_{n-h}'(K_1^{\mathbb{P}} + I_\mathcal{N})^{h-1}H^\perp\theta_\mathcal{P} + (\mathcal{B}_h - \mathcal{B}_n)' H^\perp \left(I_\mathcal{J} - P(P+R)^{-1}\right).
\end{aligned}$$

We cannot use Equation (28) directly in practice, since Z_t and $\mathcal{E}_{Z,t}^{\mathbb{P}}$ must be filtered. The filtered 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 allowing information from the past to play a role, as shown in the expression for the coefficient $\delta_{h,\mathcal{E}_\mathcal{P}}^{(n)}$ in Equation (29). Past forward rates will help predict the future bond premium whether $\theta = \Lambda_\mathcal{E} \neq 0$ 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 (29) is similar to the reduced-form specification in Equation (4).

The fact that Z_t is not Markovian does not introduce additional factors in the cross-section of yields. This holds by construction (Assumption 1). But this also holds when deriving an equivalent CM-$DTSM$ based on $\mathcal{P}_t^o$. Conditioning on $\mathcal{I}_{\mathcal{P},t}$, any yield $y_t^{(n)}$ is given by

$$y_t^{(n)} = A_{\mathcal{P},n} + B_{\mathcal{P},n}'\mathcal{P}_t^o + C_{\mathcal{P},n}'\mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}}, \tag{30}$$

with coefficients:

$$
\begin{aligned}
A_{\mathcal{P},n} &= A_n - H^{\perp} H_0 \\
B_{\mathcal{P},n} &= B_n' H^{\perp} P(P+R)^{-1} \\
C_{\mathcal{P},n} &= B_n' H^{\perp}(I_{\mathcal{J}} - P(P+R)^{-1}).
\end{aligned} \tag{31}
$$

We have that $P(P+R)^{-1}$ converges to $I_{\mathcal{J}}$ when $R \to 0$, in which case $C_{\mathcal{P},n} = 0$ even if $\theta \neq 0$.

3 Results

3.1 Summary of results

We begins with brief summary of the results. First, we find that the Markovian restriction is rejected even in a 5-factor model. The risk premium estimated based on CM dynamics exhibits higher variability and higher correlation with return-forecasting factors based on predictive regressions, and delivers more accurate forecasts of bond returns. Second, we find that a 3-factor CM model, while adding only six parameters relative to the benchmark DTSM, can raise the R^2s by as much as 25% in model-implied predictability regressions. Finally, allowing for only one factor (rank one) in the moving-average component adds only three parameters and is supported in the data. This restriction leads to essentially no difference in the results.

3.2 Data and estimation

3.2.1 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 the risk factors are Markovian. Finally, KF stands for Kalman Filter and $\mathcal{M}$ designates different structures for the covariance matrix of measurement errors R. We consider three cases:

(i) $KF0$ supposes that $\mathcal{N}$ linear combinations of the $\mathcal{J}$ yields are observed without errors, the measurement errors for the remaining $\mathcal{N}-\mathcal{J}$ linear combinations share one standard deviation parameter; (ii) $KF1$ supposes that all yields are measured with errors where $R=\sigma^2 I_{\mathcal{N}}$ and σ^2 a scalar; and (iii) $KF2$ supposes that R is an unrestricted diagonal matrix.

3.2.2 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 comparisons with the results in Cochrane and Piazzesi (2005) and Duffee (2011). For comparability, we also 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 (with the convention that $f_t^{(1)}$ is the short rate).

3.2.3 Estimation of 3-factor models

We consider several variants of 3-factor models. In these cases, we estimate the parameters $\widetilde{\Theta}^{\mathcal{P}}$—defined in Section 2.6—jointly in one step using the likelihood of $\mathcal{P}_t^o$:

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

where, $\mathcal{I}_{\mathcal{P},t-1}$ includes the history of observed portfolios (as in Theorem 1), $\mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}}$ is given by Equation (19), and the initial value $\mathcal{E}_{\mathcal{P},0}^{\mathbb{P}}$ is set to its unconditional mean. Maximum Likelihood estimation based on Equation 32 is computationally robust. We restrict the eigenvalues of $K_1^{\mathbb{Q}}$ and $K_1^{\mathbb{P}}$ within the unit circle. We use the sample average of $f_t^{(1)}$ to calibrate ρ_0, since $E[f_t^{(1)}]=\rho_0$, and we use the sample averages of $f_t^{(12)}$, $f_t^{(36)}$ and $f_t^{(60)}$ to calibrate $K_0^{\mathbb{Q}}$ (see Appendix A.4). Calibrating the level of yields makes different models more directly comparable.

3.2.4 Estimation of 5-factor models

We also consider a small number of 5-factor models for $\mathcal{P}_t^o$. We consider the CM_5^0 model where the latent factors do not share a moving-average component, but allowing different structures for the measurement errors: CM_5^0-KF, CM_5^0-$KF1$ and CM_5^0-$KF2$.

As a benchmark for most of our results, we also consider the simple case CM_5^5-$FK0$, since the identification of parameters collapses to a standard reduced-form VARMA(1,1) for $\mathcal{P}_t^o$ (see the online appendix). This is a useful benchmark, since we can compare with stylized facts documented in Section 1. As in Duffee (2011), the no-arbitrage parameters are not identified based on $\mathcal{P}_t^o$, since $\mathcal{N} = \mathcal{J}$ in this case. We proceed in two steps, as follows.

In the first step, the knife-edge case $\mathcal{J} = \mathcal{N}$ opens the road for a simpler identification where we take Σ a triangular matrix but impose that $H = \mathcal{I}_N$ (contrast with Assumption 4).[23] This is a natural assumption when the number of factors is the same as the number of observations. It is also analytically more convenient. Hence, we estimate the following parameters,

$$\Theta^{\mathcal{P}}_{\mathcal{J}=\mathcal{N}} = \{H_0, \Sigma, K_1^{\mathbb{P}}, R, \theta\}, \tag{33}$$

based on the likelihood of $\mathcal{P}_t^o$:

$$f(\mathcal{P}_t^o|\mathcal{I}_{\mathcal{P},t-1}; \Theta^{\mathcal{P}}_{\mathcal{J}=\mathcal{N}}) = -\frac{1}{2}\left(\log \Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}} + (\mathcal{P}_t^o - \mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1})'(\Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}})^{-1}(\mathcal{P}_t^o - \mathcal{E}^{\mathbb{P}}_{\mathcal{P},t-1})\right). \tag{34}$$

We use the fact that $K^{\mathbb{P}}_{0\mathcal{P}} = -K^{\mathbb{P}}_{1\mathcal{P}} H_0$ to calibrate H_0 based on sample averages. Again, this calibration makes results more comparable across models.

In a second step, we use the 3-month and 6-month yields $y_t^{(3)}$ and $y_t^{(6)}$ to estimate the $\mathbb{Q}$-parameters. Parameter estimates are not constrained by no-arbitrage in the first step. Specifically, we estimate the risk-neutral parameters $K_1^{\mathbb{Q}}$ and ρ_1, keeping the parameters $K_1^{\mathbb{P}}$, R and θ at values estimated in the first step. We calibrate ρ_0 and $K_0^{\mathbb{Q}}$ to sample averages, as in the case of 3-factor models. The second-stage estimates of $K_1^{\mathbb{Q}}$ and ρ_1 are consistent.[24] We restrict the eigenvalues of $K_1^{\mathbb{Q}}$ and $K_1^{\mathbb{P}}$ within the unit circle in all cases.

[23] Section A.2.7 of the online appendix discusses this case in detail.

[24] 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).

3.3 5-Factor Models

The first question of interest is whether non-Markovian dynamics for $\mathcal{P}_t^o$ 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 as in Duffee (2011) and similar to Adrian et al. (2013). We reduce the number of factors in Section 3.4. Note that each model has as many factors as forward rates in $\mathcal{P}_t$. Therefore, the results cannot be attributed to information in the cross-section of yields being missed by a lower-dimensional model. We drop the $KF0$ label in this section for simplicity.

3.3.1 Variance Ratios

Following Duffee (2011), we compare the variability of the bond risk premia implied from each model. This measures the gap between their respective information content. 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 maturity and horizon, 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 ranges 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 at the 1-month horizon. Consistent with Duffee (2011) and the evidence from direct regressions in Table 1, the relative contribution from past yields is highest for short-horizon returns.

3.3.2 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 $xr_{t,h}^{(n)}$ on the model bond risk premium $\widehat{brp}_{t,h}^{(n)}$:

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

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 gap between the accuracy of each model in forecasting bond returns. This gap is large. In fact, the ratios in Panel (B) are close to the ratios in Panel (A): most of the added variability of the risk premium due to the information contained in past yields translates into an improved accuracy. The relative improvements are largest 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.[25]

3.3.3 Return-Forecasting Factors

How much of the variation in the return-forecasting factor estimated via regressions in Section 1 is captured in the model-implied risk premium? Consider a regression on the risk premium implied by the term structure models,

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

where $\hat{xr}_{t,h}^{(n)} = \hat{b}_{n,h}\mathcal{R}_{t,h}$ is estimated from Equation (3). Again, we report the ratio of the R^2s obtained using $\widehat{brp}_{t,h}^{(n)}$ from the CM_5^0 model relative to that obtained using $\widehat{brp}_{t,h}^{(n)}$ from the CM_5^5 model. A ratio less than one measures the gap in matching the expected risk premium from unrestricted regressions. Figure 2 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. The ratio ranges between 65% and 95% for 12-month returns.

[25]Strictly speaking, the GDTSMs imply that $a = 0$ and $b = 1$ in Equation (35). Unreported results show that this constraint yields little change in predictability in almost every case.

3.3.4 CP's Regression

The CM_5^5 has the advantage of separating the information from current and past yields to forecast bond returns or to match the bond risk premium (in Sections 3.3.2-3.3.3, respectively). Consider CP's regression of annual bond excess returns on the current $\mathcal{P}_t^o$. We compare the population coefficients of this projection within the CM_5^0 and the CM_5^5 models, respectively. This comparison keeps the information set constant. Then, any improvement between the CM_5^0 and the CM_5^5 models must be attributed to a better description of the yields dynamics—not to difference in the information set when forming a forecast.

The more general CM_5^5 model improves forecasts even when keeping the information content fixed. Figure 3 shows the coefficients for horizons of 3, 6 and 12 months across Panels (A)-(C), respectively. We report results for bonds with maturities of 3 and 5 years, as well as estimates from direct regressions. The OLS coefficients show the familiar tent shape. The coefficients from the CM_5^5 model are close and show the expected tent shape. Consistent with results in CP, the coefficients from the CM_5^0 differ from the OLS estimates. 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 expected bond returns at longer horizons.

3.4 3-factor Models

This section focuses on models with $\mathcal{N} = 3$ factors, consistent with the stylized fact that three principal components summarize the cross-section of yields (i.e., rank$(var(\mathcal{P}_t)) = \mathcal{N} = 3$). First, we find that the simpler model CM_3^0 is rejected against any of the alternatives CM_3^1, CM_3^2 or CM_3^3. Second, we find that reducing the rank of of θ is favored in the data, based on likelihood ratio statistics or BIC model selection criteria. Our preferred CM_3^1-$KF1$ models captures most of the variability and predictability of returns in the data. One single factor drives the moving-average component and summarizes the information content in the history of yields. Finally, the results suggest that allowing for a more flexible structure of measurement errors worsens the fit in terms of returns variability and predictability.

3.4.1 Likelihood Ratio Tests

The first row of Panel (A) in Table 7 reports the number of parameters in each model. The cases with $KF0$ are presented first, followed by the cases with $KF1$ and then followed by the cases with $KF2$. The CM_3^0-$KF0$ has 19 parameters, including one standard-deviation parameter for those two combinations of yields that are measured with errors. The CM_3^3-$KF0$ has 25 parameters, allowing for six parameters for the 3×3 symmetric matrix θ. Moving to cases with $KF1$, the CM_3^0-$KF1$ has 19 parameters, the same as for the CM_3^0-$KF0$ model, since the same standard deviation parameter applies to all yields measured with errors. Models CM_3^1-$KF0$, CM_3^2-$KF0$ and CM_3^3-$KF0$ have 22, 24 and 25 parameters as we relax the rank restriction. Finally, moving to cases with $KF2$ adds four measurement error parameters. The count rises from 23, 26, 28 and to 29 between the CM_3^0 and CM_0^3 models. To highlight the gain in parsimony, the CM_5^0-$KF0$ and CM_5^5-$KF0$ models estimated in Section 3.3 have 51 and 76 parameters, respectively.

The second row of Panel (A) reports the gain in likelihood relative to CM_3^0-$KF0$. We first check whether the Markovian restriction is rejected in the data. Compare the likelihood between the CM_3^0-$KF0$ and CM_3^3-$KF0$ models: the (log-) likelihood increases by 19. The p-value of the corresponding LR test-statistics is less than 1%. The likelihood gains from relaxing the Markovian restriction for the risk factor differences are similar in the $KF1$ and $KF2$ cases, 28 and 21, respectively, and the restriction is rejected in all cases at a 1% level.

Second, we check whether reducing the rank of θ is supported in the data. The evidence in Section 1.3.3 suggested that information from the history of yields could be summarized by only one or two factors. This stylized fact would be consistent with a reduction in the rank of θ, since this reduces the number of state variables driving the bond risk premium and the dynamics of yields. Indeed, the reduction of the likelihood between the CM_3^3, CM_3^2 and CM_3^1 models are very small: only 2 and 6 in the KF_1 case and only 3 and 2 in the KF_2, respectively. The LR test-statistics fail to reject the rank restriction in every case.

The last row of Panel (A) reports the improvement of the BIC model-selection criteria relative to CM_3^0-$KF0$. This criteria leads to essentially the same conclusion: the Markovian restriction is rejected for the risk factors but, on the other hand, the restriction $\text{rank}\theta = 1$

is supported in the data.

3.4.2 Variance Ratios

One key question is whether a 3-factor CM model captures the variability and predictability of bond excess returns. In the following, we focus on annual bond returns to make results easily comparable with the existing evidence. First, we assess the variability of the model-implied risk premium. We compute the variance ratios relative to the unrestricted 5-factor CM_5^5-$KF0$ model. Panel (B) of Table 7 reports the variance ratios. A ratio close to one indicates that a good fit of the risk premium variability in the data. A ratio above one indicates that the model-implied risk premium varies excessively.

The risk-premium variability is high in the CM_3^0-$KF0$—with ratios between 1.32 and 1.50—and slightly lower in the CM_3^0-$KF0$—between 1.16 and 1.26. Moving to a Kalman filter smooths the variability further. In fact, the ratios are very close to one on all $KF1$ cases while the ratios are close to 0.80 in the $KF2$ cases. Imposing the rank restriction has essentially no effect on the results in every case.

3.4.3 Bond Returns

We also compare the R^2s from Mincer-Zarnowitz regressions of annual returns. We compute R^2 ratios relative to results obtained based on the 5-factor CM_5^5-$KF0$ model. Panel (C) of Table 7 reports the results. For the CM_3^0-$KF0$ model, the annual bond returns predictability ranges between 60 and 70 percent of that observed in the data. By contrast, the CM_3^3-$KF0$ generate bond returns predictability that is very close to the benchmark, with ratios between 86 and 93 percent. In other words, allowing for non-Markovian risk factors generate much of the predictability found in the data.

Overall the pattern are consistent with the pattern of variance ratios in Panel (B). The predictability is slightly lower when we move to $KF1$ cases–between 77 and 88 percent. This is consistent with results in Duffee (2011) for a model with three factors. Allowing for a more flexible structure of measurement errors worsens the results. The predictability ranges between 63 and 74 percent in the $KF2$ cases, relative to the benchmark. Again, reducing the rank of θ has essentially no effect on the results in every case.

3.4.4 Return-Forecasting Factors

We also check the fit of the expected-returns factor estimated via regressions in Section 1. Consider a regression of excess returns on the model risk premium obtained when $\mathcal{N} = 3$

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

which is the same as Equation 36. Figure 4 shows the ratio of the R^2s from the CM_3^3 model in the $KF0$ and $KF1$ cases, relative to that obtained from the CM_5^5-$KF0$ model. This gauges the importance of allowing for non-Markovian risk factors. The ratios are close to 1 across bond maturities and returns horizons. The ratios deteriorates significantly for the CM_3^0-$KF0$ cases and for the more general measurement error structure (unreported).

3.4.5 Out-of-Sample Yield Forecasts

We obtain rolling estimates for several three-factor models based on a rolling sample $t = 1, \ldots, T$, with T increasing each month from December 1985 until December 2015. For each month, we compute forecasts of yield changes at horizons of 1, 3, 6 and 12 months. We repeat for CM_3^3, CM_3^2 and CM_3^1 models, considering $KF1$ or $KF2$ in each case—a total of six models. We repeat for the CM_5^5-$KF0$ model, which is our benchmark once again. To aggregate forecasts across the whole sample, we compute the ratio of the out-of-sample forecast RMSE for each model relative to the RMSE from the CM_5^5-$KF0$ model. Table 8 reports the results. Cases with $KF1$ and $KF2$ are aligned in the left and right column, respectively.

The principle of parsimony appears to hold in essentially every case. Forecasts are more accurate in models with $KF1$ error structure relative to the more general $KF2$ error structure. Forecasts are more accurate in models where the moving-average component has a lower rank relative to model where this component has higher rank. Forecasts are more accurate in 3-factor models relative to the 5-factor model, except at the horizon of one-month. But the differences are small in most cases. This is consistent with results in Section 1.3.4 where the reduced-form VARMA(1,1) and VAR(1) models produced forecasts with similar accuracy even if one model is substantially more parsimonious than the other.

This also consistent with results in Adrian et al. (2013) where out-of-sample results are similar for models with four or five factors. Note that the CM_3^2-$KF2$ appears to underperform systematically, adding to the reasons for preferring the CM_3^1-$KF1$ model.

3.4.6 Measurement Errors

We also ask what are the gains derived from allowing for measurement errors while excluding a moving-average component. In the case of three-factor models, the results can be read from Table 7. Results for the CM_3^0 model with $KF0$, $KF1$ and $KF2$ are given in the first, third and seventh column, respectively. Of course, we find that allowing for more general measurement error structure generates substantial likelihood gains (Panel A). But the results also show that more general measurement errors reduce the risk premium variability substantially (Panel B) with little gains in return predictability as measured in MZ regressions (Panel C).

In the case of five-factor models, Table 9 compares the CM_5^0 model with $KF0$, $KF1$ and $KF2$ relatively to the CM_5^5-$KF0$. The likelihood gains associated with more general measurement errors are much smaller with five risk factors than with three factors, as expected. But, similar to the result from 3-factor models, we observe a reduction in the variability and predictive content of model risk premium (Panel B-C). We find one exception across all results in Tables 7- 9: the return predictability for the CM_5^5-$KF2$ model are very close to the benchmark.[26]

4 Yield Decomposition

Estimates of the risk premium from CM models that include a moving-average component are close to estimates from predictability regressions. In turn, these CM models provide better forecasts of bond returns. This implies that CM models provide a more accurate decomposition of yields between an expectation component and a risk component. The results also suggest that the rank of the common moving-average component has no effect on the decomposition of yields. Indeed, Figure 5 compares the 5-year term premium from

[26]The increased performance of CM_5^5-$KF2$ in matching MZ regression R^2s hide much higher correlations between lagged residuals–much higher than even the simplest VAR(1) (see Panel B of Table 4).

models CM_3^3, CM_3^2 and CM_3^1 in the $KF1$ case, showing that there is essentially no difference in the model-implied risk premium.

Focusing on the parsimonious CM_3^1-$KF1$ model, Figure 6 compares the decomposition of the 5-year yields between the CM_3^1-$KF1$ and CM_3^0-$KF0$. Panel (A) shows the decomposition from the CM_3^0-$KF0$ model. Panel (B) shows the decomposition from the CM_3^3-$KF1$ specification. Panel (C) isolates the term premium from each model.

The term premium is substantially more variable and cyclical in the CM_3^3-$KF1$ model. Allowing a role for the history of yields attributes a much greater share of variations to the risk premium, and produces smoother estimates of interest rate expectations. For instance, the term premium estimate is lower early in recessions, while the expectation component falls by less and recovers faster as the economy approaches the recovery. The episode between 2001 and 2005 provides a powerful case in point. Using the information from past yields produces a larger decline of term premium estimates early in 2001, by as much as 1.5 percent and the difference persists throughout the recession. Hence, the downward adjustment in expectations is far smaller in the benchmark model around the initial phase of the recessions.

Strikingly, this difference is also apparent during the "conundrum" episode in 2004-2005. The term premium from the CM_3^3-$KF1$ model falls steadily as the Federal Reserve raises short-term interest rates, but the risk premium from the CM_3^0-$KF0$ remains essentially unchanged 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.[27]

5 Conclusion

We revisit and extend the evidence in Cochrane and Piazzesi (2005) and Duffee (2011). As noted by Cochrane and Piazzesi (2005) and Duffee (2011), among others, the additional information content in the history or yields is inconsistent with the standard GDTSM

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

specification. This paper proposes a simple, parsimonious and intuitive reconciliation, where we identify two separate channels introducing a role for a moving-average in the dynamics for yields. First, the latent risk factors themselves 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.

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 non-Markovian: is it because the true state exhibits 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. Finally, the CM representation based on the steady-state Kalman filter should be useful in several applications, since it combines the Kalman filter and Maximum Likelihood estimation in a single convenient step.

References

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

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.

Chernov, M. and P. Mueller (2011). The term structure of inflation expectations. *Journal of Financial Economics 106*(2), 367–394.

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 (2000). Specification analysis of affine term structure models. *The Journal of Finance 55*, 1943–1978.

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.

Diebold, F. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*, 337–364.

Doornik, J. (1996). Testing vector error autocorrelation and heteroscedasticity. Working paper, Oxford University.

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

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 (2014). Non-markov Gaussian term structure models: The case of inflation. *Review of Finance 18*(5), 1953–2001.

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.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica 46*(6), pp. 1303–1310.

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.

Hansen, L. and T. J. Sargent (2014). *Recursive models of dynamic linear economies*. Princeton University Press.

Hayashi, F. (2001). *Econometrics*. Princeton University Press.

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 11*(4), 581–609.

Joslin, S., M. Priebsch, and K. J. Singleton (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance 69*(3), 1197–1233.

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.

Le, A., K. J. Singleton, and Q. Dai (2010). Discrete-time affine $\mathbb{Q}$ term structure models with generalized market prices of risk. *Review of Financial Studies 23*(5), 2184–2227.

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

Lütkephol, H. (1984). Linear transformations of vector ARMA processes. *Journal of Econometrics 26*(3), 283 – 293.

Lütkephol, H. (2006). *New Introduction to Multiple Time Series Analysis*. Springer.

Maddala, G. (2001). *Introduction to Econometrics* (3rd ed.). Wiley.

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.

A Appendix

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

The yields 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 $E[Y_{t+h}|Y_t, \dots,] \equiv E_t[Y_{t+h}] = C_h \sum_{i=0}^{\infty} \Theta_i Y_{t-i}$ 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 $i = 0, \dots, \infty$ 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

Suppose that the conditional distribution of $Z_t \in \mathbb{R}^{\mathcal{N}}$ under $\mathbb{Q}$ is Gaussian with constant variance $\Sigma\Sigma'$, possibly conditional on Z_{t-1} and its past, or some other factor $Z_{2,t}$. Suppose that the cross-section of n-period yields for $n \geq 1$, can be expressed as a linear function of Z_t,

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

with the short rate given by

$$r_t = y_t^{(1)} = \rho_0 + \rho_1 Z_t. \tag{42}$$

In the absence of arbitrage opportunity, we have that

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

For $n = 2$:

$$\begin{aligned} y_t^{(2)} &= -\frac{1}{2} \log \left(\exp\left(-\rho_0 - \rho_1' Z_t\right) E_t^{\mathbb{Q}} \left[\exp\left(-\rho_0 - \rho_1' Z_{t+1}\right) \right] \right) \\ &= -\frac{1}{2} \log \left(\exp \left(-2\rho_0 - \rho_1' Z_t - \rho_1' \mathcal{E}_{Z,t}^{\mathbb{Q}} + \frac{\rho_1' \Sigma\Sigma' \rho_1}{2} \right) \right) \\ &= -\frac{1}{2} \left(-2\rho_0 + \frac{\rho_1' \Sigma\Sigma' \rho_1}{2} - \rho_1' Z_t - \rho_1' \mathcal{E}_{Z,t}^{\mathbb{Q}} \right), \end{aligned}$$

where the conditional mean $\mathcal{E}^{\mathbb{Q}}_{Z,t}$ must be a linear function of Z_t for Equation (41) to hold, say $\mathcal{E}^{\mathbb{Q}}_{Z,t} = K^{\mathbb{Q}}_0 + (K^{\mathbb{Q}}_1 + I_{\mathcal{N}})Z_t$ as in the proposition:

$$\begin{aligned}
y^{(2)}_t &= -\frac{1}{2}\left(-2\rho_0 + \frac{\rho_1'\Sigma\Sigma'\rho_1}{2}\right) - \frac{1}{2}\left(-\rho_1' Z_t - \rho_1' K^{\mathbb{Q}}_0 + \rho_1'(K^{\mathbb{Q}}_1 + I_{\mathcal{N}})Z_t\right) \\
&= -\frac{1}{2}\left(-2\rho_0 + \frac{\rho_1'\Sigma\Sigma'\rho_1}{2} + \rho_1' K^{\mathbb{Q}}_0\right) - \frac{1}{2}\left(-\rho_1' - \rho_1'(K^{\mathbb{Q}}_1 + I_{\mathcal{N}})\right) Z_t \\
&= A_2 + B_2' Z_t.
\end{aligned}$$

It is easy to check that the same argument holds for $n > 2$. Therefore, we can write

$$Z_t = \mathcal{E}^{\mathbb{Q}}_{Z,t-1} + \Sigma\epsilon^{\mathbb{Q}}_t = K^{\mathbb{Q}}_0 + (K^{\mathbb{Q}}_1 + I_{\mathcal{N}})Z_{t-1} + \Sigma\epsilon^{\mathbb{Q}}_t, \tag{43}$$

where $\epsilon^{\mathbb{Q}}_t$ is a standard Gaussian innovation, as required.

To show necessity, suppose that Z_t has the $\mathbb{Q}$-dynamics given in the proposition, it follows from standard results (Le, Singleton, and Dai, 2010) that the no-arbitrage price at time-t of zero-coupon bonds with n-period to maturity $D^{(n)}_t$ is given by

$$\begin{aligned}
D^{(n)}_t &= E^{\mathbb{Q}}_t[e^{-\sum_{i=0}^{n-1} r_{t+i}}] \\
&= e^{\mathcal{A}_n + \mathcal{B}_n' Z_t},
\end{aligned} \tag{44}$$

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

$$\begin{aligned}
\mathcal{A}_{n+1} - \mathcal{A}_n &= K^{\mathbb{Q}\prime}_0 \mathcal{B}_n + \frac{1}{2}\mathcal{B}_n'\Sigma\Sigma'\mathcal{B}_n - \rho_0 \\
\mathcal{B}_{n+1} - \mathcal{B}_n &= K^{\mathbb{Q}\prime}_1 \mathcal{B}_n - \rho_1,
\end{aligned} \tag{45}$$

and zero-coupon yields satisfy $y^{(n)}_t = -\frac{1}{n}D^{(n)}_t = A_n + B_n' Z_t$ as required, with coefficients given by $A_n = -\mathcal{A}_n/n$ and $B_n = -\mathcal{B}_n/n$.

Proposition 2

Let $Z_t \in \mathbb{R}^{\mathcal{N}}$ with Markovian $\mathbb{Q}$-dynamics as in Proposition 1. Define $\epsilon^{\mathbb{P}}_t$ and $\mathcal{E}^{\mathbb{P}}_{Z,t-1}$:

$$Z_t = \mathcal{E}^{\mathbb{P}}_{Z,t-1} + \Sigma\epsilon^{\mathbb{P}}_t, \tag{46}$$

with $\epsilon^{\mathbb{P}}_t$ a standard Gaussian white noise. Consider the change of measure M_{t+1} given by

$$M_{t+1} \equiv \exp\left(-\frac{\Lambda_t'\Sigma\Sigma'\Lambda_t}{2} - \Lambda_t'\Sigma\epsilon^{\mathbb{P}}_{t+1}\right). \tag{47}$$

Then,

$$E^{\mathbb{Q}}_t[\exp(u'\epsilon^{\mathbb{P}}_{t+1})] = E_t[M_{t+1}\exp(u'\epsilon^{\mathbb{P}}_{t+1})] = \exp(\frac{u'u}{2} - u'\Sigma'\Lambda_t), \tag{48}$$

which implies that $\epsilon^{\mathbb{P}}_{t+1} = -\Sigma'\Lambda_t + \epsilon^{\mathbb{Q}}_{t+1}$ and, using Equation 11, that

$$\begin{aligned}
\Delta Z_{t+1} &= K^{\mathbb{Q}}_0 + K^{\mathbb{Q}}_1 Z_t + \Sigma(\epsilon^{\mathbb{P}}_{t+1} + \Sigma'\Lambda_t) \\
&= K^{\mathbb{Q}}_0 + K^{\mathbb{Q}}_1 Z_t + \Sigma\Sigma'\Lambda_t + \Sigma\epsilon^{\mathbb{P}}_{t+1}.
\end{aligned} \tag{49}$$

If the prices of risk Λ_t are given by

$$\Lambda_t \equiv (\Sigma\Sigma')^{-1}\left(\Lambda_0 + \Lambda_Z Z_t + \Lambda_{\mathcal{E}}\mathcal{E}^{\mathbb{P}}_{Z,t-1}\right), \tag{50}$$

and taking the conditional expectation in Equation 49, it follows that

$$\begin{aligned}\mathcal{E}^{\mathbb{P}}_{Z,t} &= K^{\mathbb{Q}}_0 + (K^{\mathbb{Q}}_1 + I_{\mathcal{N}})Z_t + \Lambda_0 + \Lambda_Z Z_t + \Lambda_{\mathcal{E}}\mathcal{E}^{\mathbb{P}}_{Z,t-1}\\ &= (K^{\mathbb{Q}}_0 + \Lambda_0) + (K^{\mathbb{Q}}_1 + I_{\mathcal{N}} + \Lambda_Z)Z_t + \Lambda_{\mathcal{E}}\mathcal{E}^{\mathbb{P}}_{Z,t-1}\\ &= (K^{\mathbb{Q}}_0 + \Lambda_0) + (K^{\mathbb{Q}}_1 + I_{\mathcal{N}} + \Lambda_Z)(\mathcal{E}^{\mathbb{P}}_{Z,t-1} + \Sigma\epsilon^{\mathbb{P}}_t) + \Lambda_{\mathcal{E}}\mathcal{E}^{\mathbb{P}}_{Z,t-1}\\ &= (K^{\mathbb{Q}}_0 + \Lambda_0) + (K^{\mathbb{Q}}_1 + I_{\mathcal{N}} + \Lambda_Z + \Lambda_{\mathcal{E}})\mathcal{E}^{\mathbb{P}}_{Z,t-1} + (K^{\mathbb{Q}}_1 + I_{\mathcal{N}} + \Lambda_Z)\Sigma\epsilon^{\mathbb{P}}_t, \end{aligned} \tag{51}$$

where we used $Z_t = \mathcal{E}^{\mathbb{P}}_{Z,t-1} + \Sigma\epsilon^{\mathbb{P}}_t$. This yields the dynamics in Equation 14 from Assumption 2 with the unique mapping between parameters of the $\mathbb{P}$- and $\mathbb{Q}$-dynamics as given in Proposition 2.

Proposition 3

The excess returns on investing on a n maturity bond, over a holding period h is

$$\begin{aligned} xr^{(n)}_{t,h} &= -(n-h)y^{(n)}_{t+h} + ny^{(n)}_t - hy^{(h)}_t\\ &= \mathcal{B}'_{n-h}Z_{t+h} + (\mathcal{B}_h - \mathcal{B}_n)' Z_t + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n. \end{aligned}$$

Therefore, the risk premium from holding an n-period bond between t and $t+h$ is

$$\begin{aligned} brp^{(n)}_{t,h} &\equiv E^{\mathbb{P}}[xr^{(n)}_{t,h}|\mathcal{I}_{Z,t}],\\ &= \mathcal{B}'_{n-h}E[Z_{t+h}|\mathcal{I}_{Z,t}] + (\mathcal{B}_h - \mathcal{B}_n)' Z_t + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n,\\ &= \mathcal{B}'_{n-h}E[\mathcal{E}^{\mathbb{P}}_{Z,t+h-1}|\mathcal{I}_{Z,t}] + (\mathcal{B}_h - \mathcal{B}_n)' Z_t + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n,\\ &= \mathcal{B}'_{n-h}\left(E[Z] + (K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1}\left(\mathcal{E}^{\mathbb{P}}_{Z,t} - E[Z]\right)\right) + (\mathcal{B}_h - \mathcal{B}_n)' Z_t + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n.\\ &\equiv \omega^{(n)}_h + \beta^{(n)\prime}_h\mathcal{E}^{\mathbb{P}}_{Z,t} + (\mathcal{B}_h - \mathcal{B}_n)' Z_t\\ &= \omega^{(n)}_h + \beta^{(n)\prime}_h\left[K^{\mathbb{P}}_0 + \left(K^{\mathbb{P}}_1 + I_{\mathcal{N}} - \theta\right)Z_t + \theta\mathcal{E}^{\mathbb{P}}_{Z,t-1}\right] + (\mathcal{B}_h - \mathcal{B}_n)' Z_t\\ &\equiv \delta^{(n)}_{h,0} + \delta^{(n)\prime}_{h,Z}Z_t + \delta^{(n)\prime}_{h,\mathcal{E}}\mathcal{E}^{\mathbb{P}}_{Z,t-1}, \end{aligned}$$

where

$$\begin{aligned} E[Z] &\equiv -(K^{\mathbb{P}}_1)^{-1}K^{\mathbb{P}}_0,\\ \omega^{(n)}_h &\equiv \mathcal{B}'_{n-h}E[Z] - \mathcal{B}'_{n-h}(K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1}E[Z] + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n,\\ \beta^{(n)}_h &\equiv (K^{\mathbb{P}\prime}_1 + I_{\mathcal{N}})^{h-1}\mathcal{B}_{n-h},\\ \delta^{(n)}_{h,0} &= \mathcal{B}'_{n-h}\left[(K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1}(K^{\mathbb{P}}_1)^{-1} + (K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1} - (K^{\mathbb{P}}_1)^{-1}\right]K^{\mathbb{P}}_0 + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n,\\ \delta^{(n)\prime}_{h,Z} &= \mathcal{B}'_{n-h}(K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1}\left(K^{\mathbb{P}}_1 + I_{\mathcal{N}} - \theta\right) + (\mathcal{B}_h - \mathcal{B}_n)',\\ \delta^{(n)\prime}_{h,\mathcal{E}} &= \mathcal{B}'_{n-h}(K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1}\theta. \end{aligned}$$

where the constant $\delta^{(n)}_{h,0\mathcal{P}}$ simplifies to

$$\begin{aligned} \delta^{(n)}_{h,0} &= \mathcal{B}'_{n-h}\left[(K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1}(K^{\mathbb{P}}_1)^{-1} + (K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1} - (K^{\mathbb{P}}_1)^{-1}\right]K^{\mathbb{P}}_0 + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n,\\ &= \mathcal{B}'_{n-h}\left[(K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1} + (K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1}K^{\mathbb{P}}_1 - I_{\mathcal{N}}\right](K^{\mathbb{P}}_1)^{-1}K^{\mathbb{P}}_0 + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n,\\ &= \mathcal{B}'_{n-h}\left[(K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1} + (K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h-1}\left(K^{\mathbb{P}}_1 + I_{\mathcal{N}} - I_{\mathcal{N}}\right) - I_{\mathcal{N}}\right](K^{\mathbb{P}}_1)^{-1}K^{\mathbb{P}}_0 + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n,\\ &= \mathcal{B}'_{n-h}\left[(K^{\mathbb{P}}_1 + I_{\mathcal{N}})^{h} - I_{\mathcal{N}}\right](K^{\mathbb{P}}_1)^{-1}K^{\mathbb{P}}_0 + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n, \end{aligned}$$

Similarly, using $\theta = \Lambda_{\mathcal{E}}$, we have

$$\begin{aligned}
\delta_{h,\mathcal{E}}^{(n)\prime} &= \mathcal{B}_{n-h}^{\prime}(K_1^{\mathbb{P}} + I_{\mathcal{N}})^{h-1}\Lambda_{\mathcal{E}}, \\
\delta_{h,Z}^{(n)\prime} &= \mathcal{B}_{n-h}^{\prime}\left((K_1^{\mathbb{P}} + I_{\mathcal{N}})^{h} - (K_1^{\mathbb{P}} + I_{\mathcal{N}})^{h-1}\Lambda_{\mathcal{E}}\right) + (\mathcal{B}_h - \mathcal{B}_n)^{\prime}.
\end{aligned}$$

Conditioning on $\mathcal{I}_{\mathcal{P},t}$, the information generated by $\mathcal{P}_t^o$, the expected value of the bond risk premium is given by

$$\begin{aligned}
\widehat{brp}_{t,h}^{(n)} &\equiv E^{\mathbb{P}}[brp_{t,h}^{(n)}|\mathcal{I}_{\mathcal{P},t}], \\
&= E^{\mathbb{P}}[\omega_h^{(n)} + \beta_h^{(n)\prime}\mathcal{E}_{Z,t}^{\mathbb{P}} + (\mathcal{B}_h - \mathcal{B}_n)^{\prime} Z_t|\mathcal{I}_{\mathcal{P},t}] \\
&= \omega_h^{(n)} + \beta_h^{(n)\prime}E^{\mathbb{P}}\left[\mathcal{E}_{Z,t}^{\mathbb{P}}|\mathcal{I}_{\mathcal{P},t}\right] + (\mathcal{B}_h - \mathcal{B}_n)^{\prime} E^{\mathbb{P}}\left[Z_t|\mathcal{I}_{\mathcal{P},t}\right]
\end{aligned}$$

Given that $\mathcal{P}_t = H_0 + H^{\prime}Z_t$ and $\mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} \equiv E_t^{\mathbb{P}}[\mathcal{P}_{t+1}|\mathcal{I}_{Z,t}] = H_0 + H^{\prime}\mathcal{E}_{Z,t}^{\mathbb{P}}$, and using $H(H^{\perp})^{\top} = I_{\mathcal{N}}$, we have

$$Z_t = H^{\perp}\left(\mathcal{P}_t - H_0\right), \quad \mathcal{E}_{Z,t}^{\mathbb{P}} = H^{\perp}\left(\mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} - H_0\right)$$

and therefore:

$$\begin{aligned}
\widehat{brp}_{t,h}^{(n)} = {} & \omega_h^{(n)} - \beta_h^{(n)\prime}H^{\perp}H_0 - (\mathcal{B}_h - \mathcal{B}_n)^{\prime} H^{\perp}H_0 \\
& + \beta_h^{(n)\prime}H^{\perp}E^{\mathbb{P}}\left[\mathcal{E}_{\mathcal{P},t}^{\mathbb{P}}|\mathcal{I}_{\mathcal{P},t}\right] + (\mathcal{B}_h - \mathcal{B}_n)^{\prime} H^{\perp}E^{\mathbb{P}}\left[\mathcal{P}_t|\mathcal{I}_{\mathcal{P},t}\right].
\end{aligned}$$

Using results from the steady-state representation of the Kalman filter in Appendix A.3, we have

$$\begin{aligned}
E^{\mathbb{P}}\left[\mathcal{E}_{\mathcal{P},t}^{\mathbb{P}}|\mathcal{I}_{\mathcal{P},t}\right] &= E^{\mathbb{P}}\left[\mathcal{P}_{t+1}|\mathcal{I}_{\mathcal{P},t}\right] = E^{\mathbb{P}}\left[\mathcal{P}_{t+1}^o|\mathcal{I}_{\mathcal{P},t}\right] \\
E^{\mathbb{P}}\left[\mathcal{P}_t|\mathcal{I}_{\mathcal{P},t}\right] &= E^{\mathbb{P}}\left[\mathcal{P}_t|\mathcal{I}_{\mathcal{P},t-1}\right] + P(P+R)^{-1}\left(\mathcal{P}_t^o - E^{\mathbb{P}}\left[\mathcal{P}_t^o|\mathcal{I}_{\mathcal{P},t-1}\right]\right) \\
&= E^{\mathbb{P}}\left[\mathcal{P}_t^o|\mathcal{I}_{\mathcal{P},t-1}\right] + P(P+R)^{-1}\left(\mathcal{P}_t^o - E^{\mathbb{P}}\left[\mathcal{P}_t^o|\mathcal{I}_{\mathcal{P},t-1}\right]\right)
\end{aligned}$$

Define $\mathcal{E}_{\mathcal{P}^o,t-1}^{\mathbb{P}} = E^{\mathbb{P}}\left[\mathcal{P}_t^o|\mathcal{I}_{\mathcal{P},t-1}\right]$, the bond risk premium is

$$\begin{aligned}
\widehat{brp}_{t,h}^{(n)} &= \omega_h^{(n)} - \beta_h^{(n)\prime}H^{\perp}H_0 - (\mathcal{B}_h - \mathcal{B}_n)^{\prime} H^{\perp}H_0 + \beta_h^{(n)\prime}H^{\perp}E^{\mathbb{P}}\left[\mathcal{P}_{t+1}^o|\mathcal{I}_{\mathcal{P},t}\right] \\
&\quad + (\mathcal{B}_h - \mathcal{B}_n)^{\prime} H^{\perp}\left(E^{\mathbb{P}}\left[\mathcal{P}_t^o|\mathcal{I}_{\mathcal{P},t-1}\right] + P(P+R)^{-1}\left(\mathcal{P}_t^o - E^{\mathbb{P}}\left[\mathcal{P}_t^o|\mathcal{I}_{\mathcal{P},t-1}\right]\right)\right) \\
&= \omega_h^{(n)} - \beta_h^{(n)\prime}H^{\perp}H_0 - (\mathcal{B}_h - \mathcal{B}_n)^{\prime} H^{\perp}H_0 + \beta_h^{(n)\prime}H^{\perp}\left[K_{0\mathcal{P}^o}^{\mathbb{P}} + \left(K_{1\mathcal{P}}^{\mathbb{P}} + I_{\mathcal{J}} - \theta_{\mathcal{P}}\right)\mathcal{P}_t^o + \theta_{\mathcal{P}}\mathcal{E}_{\mathcal{P}^o,t-1}^{\mathbb{P}}\right] \\
&\quad + (\mathcal{B}_h - \mathcal{B}_n)^{\prime} H^{\perp}\left(\mathcal{E}_{\mathcal{P}^o,t-1}^{\mathbb{P}} + P(P+R)^{-1}\left(\mathcal{P}_t^o - \mathcal{E}_{\mathcal{P}^o,t-1}^{\mathbb{P}}\right)\right)
\end{aligned}$$

Finally, grouping terms leads to

$$\begin{aligned}
\widehat{brp}_{t,h}^{(n)} &= \omega_h^{(n)} - \beta_h^{(n)\prime}H^{\perp}H_0 - (\mathcal{B}_h - \mathcal{B}_n)^{\prime} H^{\perp}H_0 + \beta_h^{(n)\prime}H^{\perp}K_{0\mathcal{P}^o}^{\mathbb{P}} \\
&\quad + \left(\beta_h^{(n)\prime}H^{\perp}\left(K_{1\mathcal{P}}^{\mathbb{P}} + I_{\mathcal{J}} - \theta_{\mathcal{P}}\right) + (\mathcal{B}_h - \mathcal{B}_n)^{\prime} H^{\perp}P(P+R)^{-1}\right)\mathcal{P}_t^o \\
&\quad + \left(\beta_h^{(n)\prime}H^{\perp}\theta_{\mathcal{P}} + (\mathcal{B}_h - \mathcal{B}_n)^{\prime} H^{\perp} - (\mathcal{B}_h - \mathcal{B}_n)^{\prime} H^{\perp}P(P+R)^{-1}\right)\mathcal{E}_{\mathcal{P}^o,t-1}^{\mathbb{P}},
\end{aligned}$$

which shows the desired result:

$$\widehat{brp}_{t,h}^{(n)} \equiv \delta_{h,0\mathcal{P}}^{(n)} + \delta_{h,\mathcal{P}}^{(n)\prime}\mathcal{P}_t^o + \delta_{h,\mathcal{E}_{\mathcal{P}}}^{(n)\prime}\mathcal{E}_{\mathcal{P}^o,t-1}^{\mathbb{P}},$$

with

$$
\begin{aligned}
\delta_{h,0\mathcal{P}}^{(n)} &\equiv \omega_h^{(n)} - \beta_h^{(n)\prime} H^{\perp} H_0 - (\mathcal{B}_h - \mathcal{B}_n)' H^{\perp} H_0 + \beta_h^{(n)\prime} H^{\perp} K_{0\mathcal{P}^o}^{\mathbb{P}} \\
&= \mathcal{B}_{n-h}' \left((K_1^{\mathbb{P}} + I_{\mathcal{N}})^h - I_{\mathcal{N}} \right) (K_1^{\mathbb{P}})^{-1} K_0^{\mathbb{P}} + \mathcal{A}_{n-h} + \mathcal{A}_h - \mathcal{A}_n \\
&\quad - \left(\mathcal{B}_{n-h}' (K_1^{\mathbb{P}} + I_{\mathcal{N}})^{h-1} K_1^{\mathbb{P}} + (\mathcal{B}_h - \mathcal{B}_n)' \right) H^{\perp} H_0 \\
\delta_{h,\mathcal{P}}^{(n)\prime} &\equiv \beta_h^{(n)\prime} H^{\perp} \left(K_{1\mathcal{P}}^{\mathbb{P}} + I_{\mathcal{J}} - \theta_{\mathcal{P}} \right) + (\mathcal{B}_h - \mathcal{B}_n)' H^{\perp} P (P+R)^{-1} \\
&= \mathcal{B}_{n-h}' (K_1^{\mathbb{P}} + I_{\mathcal{N}})^{h-1} H^{\perp} \left(K_{1\mathcal{P}}^{\mathbb{P}} + I_{\mathcal{J}} - \theta_{\mathcal{P}} \right) + (\mathcal{B}_h - \mathcal{B}_n)' H^{\perp} P (P+R)^{-1} \\
\delta_{h,\mathcal{E}}^{(n)\prime} &\equiv \mathcal{B}_{n-h}' (K_1^{\mathbb{P}} + I_{\mathcal{N}})^{h-1} H^{\perp} \theta_{\mathcal{P}} + (\mathcal{B}_h - \mathcal{B}_n)' H^{\perp} \left(I_{\mathcal{J}} - P (P+R)^{-1} \right).
\end{aligned}
$$

A.3 Conditional Mean Representation of the Kalman Filter

Theorem 1

In the first part the proof, the conditional expectations are computed with respect to the information set generated by Z_t, i.e., $E_t[\cdot] = E[\cdot | Z_t, Z_{t-1}, \ldots, Z_1]$. 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 $\mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} \equiv E_t^{\mathbb{P}}[\mathcal{P}_{t+1}] = H_0 + H' \mathcal{E}_{Z,t}^{\mathbb{P}}$. Then, the unobserved portfolios $\mathcal{P}_t$ have the following dynamics:

$$
\begin{aligned}
\mathcal{P}_{t+1} &= \mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} + H' \Sigma \epsilon_{t+1}^{\mathbb{P}}, \\
\mathcal{E}_{\mathcal{P},t+1}^{\mathbb{P}} &= K_{0\mathcal{P}}^{\mathbb{P}} + (K_{1\mathcal{P}}^{\mathbb{P}} + I_{\mathcal{J}}) \mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} + H' \Sigma_{\mathcal{E}_Z} \epsilon_{t+1}^{\mathbb{P}},
\end{aligned} \tag{52}
$$

with

$$
\begin{aligned}
K_{0\mathcal{P}}^{\mathbb{P}} &= H' K_0^{\mathbb{P}} - K_{1\mathcal{P}}^{\mathbb{P}} H_0 \\
K_{1\mathcal{P}}^{\mathbb{P}} &= H' (K_1^{\mathbb{P}} + I_{\mathcal{N}}) H^{\perp} - I_{\mathcal{J}}.
\end{aligned}
$$

Indeed from $\mathcal{P}_t = H_0 + H' Z_t$:

$$
\begin{aligned}
\mathcal{P}_t &= H_0 + H' (\mathcal{E}_{Z,t-1}^{\mathbb{P}} + \Sigma \epsilon_t^{\mathbb{P}}) \\
&= \mathcal{E}_{\mathcal{P},t-1}^{\mathbb{P}} + H' \Sigma \epsilon_t^{\mathbb{P}},
\end{aligned}
$$

while from $\mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} = H_0 + H' \mathcal{E}_{Z,t}^{\mathbb{P}}$:

$$
\begin{aligned}
\mathcal{E}_{\mathcal{P},t+1}^{\mathbb{P}} &= H_0 + H' \mathcal{E}_{Z,t+1}^{\mathbb{P}} \\
&= H_0 + H' (K_0^{\mathbb{P}} + (K_1^{\mathbb{P}} + I_{\mathcal{N}}) \mathcal{E}_{Z,t}^{\mathbb{P}} + \Sigma_{\mathcal{E}_Z} \epsilon_{t+1}^{\mathbb{P}}) \\
&= H_0 + H' K_0^{\mathbb{P}} + H' (K_1^{\mathbb{P}} + I_{\mathcal{N}}) \mathcal{E}_{Z,t}^{\mathbb{P}} + H' \Sigma_{\mathcal{E}_Z} \epsilon_{t+1}^{\mathbb{P}} \\
&= H_0 + H' K_0^{\mathbb{P}} + H' (K_1^{\mathbb{P}} + I_{\mathcal{N}}) H^{\perp} H' \mathcal{E}_{Z,t}^{\mathbb{P}} + H' \Sigma_{\mathcal{E}_Z} \epsilon_{t+1}^{\mathbb{P}} \\
&= H_0 + H' K_0^{\mathbb{P}} + H' (K_1^{\mathbb{P}} + I_{\mathcal{N}}) H^{\perp} (\mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} - H_0) + H' \Sigma_{\mathcal{E}_Z} \epsilon_{t+1}^{\mathbb{P}} \\
&= H_0 + H' K_0^{\mathbb{P}} - H' (K_1^{\mathbb{P}} + I_{\mathcal{N}}) H^{\perp} H_0 + H' (K_1^{\mathbb{P}} + I_{\mathcal{N}}) H^{\perp} \mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} + H' \Sigma_{\mathcal{E}_Z} \epsilon_{t+1}^{\mathbb{P}} \\
&\equiv H_0 + H' K_0^{\mathbb{P}} - (K_{1\mathcal{P}}^{\mathbb{P}} + I_{\mathcal{J}}) H_0 + (K_{1\mathcal{P}}^{\mathbb{P}} + I_{\mathcal{J}}) \mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} + H' \Sigma_{\mathcal{E}_Z} \epsilon_{t+1}^{\mathbb{P}} \\
&= H' K_0^{\mathbb{P}} - K_{1\mathcal{P}}^{\mathbb{P}} H_0 + (K_{1\mathcal{P}}^{\mathbb{P}} + I_{\mathcal{J}}) \mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} + H' \Sigma_{\mathcal{E}_Z} \epsilon_{t+1}^{\mathbb{P}} \\
&\equiv K_{0\mathcal{P}}^{\mathbb{P}} + (K_{1\mathcal{P}}^{\mathbb{P}} + I_{\mathcal{J}}) \mathcal{E}_{\mathcal{P},t}^{\mathbb{P}} + H' \Sigma_{\mathcal{E}_Z} \epsilon_{t+1}^{\mathbb{P}}.
\end{aligned}
$$

The observed portfolios are measured with errors,

$$
\mathcal{P}_t^o = \mathcal{P}_t + w_t, \tag{53}
$$

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

$$\begin{aligned}\mathcal{X}_{t+1} &= \mu_X + F_X\mathcal{X}_t + Q_X^{1/2}\epsilon_t^{\mathbb{P}} \\ \mathcal{P}_t^o &= H_X'\mathcal{X}_t + w_t,\end{aligned} \tag{54}$$

where

$$\mu_X = \begin{pmatrix} 0 \\ K_{0\mathcal{P}}^{\mathbb{P}} \end{pmatrix} \quad F_X = \begin{bmatrix} 0 & I_{\mathcal{J}} \\ 0 & K_{1\mathcal{P}}^{\mathbb{P}} + I_{\mathcal{J}} \end{bmatrix} \quad H_X = \begin{bmatrix} I_{\mathcal{J}} \\ 0 \end{bmatrix} \quad Q_X^{1/2} = \begin{bmatrix} H'\Sigma \\ H'\Sigma_{\mathcal{E}} \end{bmatrix} \tag{55}$$

Given this state-space representation, the second part of the proof follows the classic Kalman filtering procedure (See Hamilton 1994, Chapter 13, pp. 377-381). In the remainder of the proof, the conditional expectations are computed with respect to the information set generated by $\mathcal{P}_t^o$, i.e., $E_t[\cdot] = E[\cdot|\mathcal{P}_t^o, \mathcal{P}_{t-1}^o, \ldots, \mathcal{P}_1^o]$. This slight abuse of notation makes for better readability. To be clear, the Kalman filter forecast $E_t^{\mathbb{P}}[\mathcal{P}_{t+1}^o]$ is conditional on the observations $\{\mathcal{P}_t^o, \mathcal{P}_{t-1}^o, \ldots, \mathcal{P}_1^o\}$ and some initial value $\mathcal{E}^{\mathbb{P}}_{\mathcal{P},0}$. From standard results, the filtered estimates are given by:

$$E_t^{\mathbb{P}}[\mathcal{P}_{t+1}^o] = H_X'\left[\mu_X + F_X\left\{E_{t-1}^{\mathbb{P}}[\mathcal{X}_t] + P_{t|t-1}H_X(H_X'P_{t|t-1}H_X + R)^{-1}\left(\mathcal{P}_t^o - E_{t-1}^{\mathbb{P}}[\mathcal{P}_t^o]\right)\right\}\right], \tag{56}$$

with

$$P_{t+1|t} \equiv E_t^{\mathbb{P}}\left[(\mathcal{X}_{t+1} - E_t^{\mathbb{P}}[\mathcal{X}_{t+1}])(\mathcal{X}_{t+1} - E_t^{\mathbb{P}}[\mathcal{X}_{t+1}])'\right]$$

Given that $\mathcal{X}_t' \equiv (\mathcal{P}_t' \ \mathcal{E}^{\mathbb{P}\prime}_{\mathcal{P},t})$, we use the following block notation for $P_{t+1|t}$:

$$P_{t+1|t} = \begin{bmatrix} P_{t+1|t}^{11} & \left(P_{t+1|t}^{21}\right)' \\ P_{t+1|t}^{21} & P_{t+1|t}^{22} \end{bmatrix},$$

where

$$\begin{aligned}P_{t+1|t}^{11} &\equiv E_t^{\mathbb{P}}\left[(\mathcal{P}_{t+1} - E_t^{\mathbb{P}}[\mathcal{P}_{t+1}])(\mathcal{P}_{t+1} - E_t^{\mathbb{P}}[\mathcal{P}_{t+1}])'\right] \\ P_{t+1|t}^{21} &\equiv E_t^{\mathbb{P}}\left[(\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t+1} - E_t^{\mathbb{P}}[\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t+1}])(\mathcal{P}_{t+1} - E_t^{\mathbb{P}}[\mathcal{P}_{t+1}])'\right] \\ P_{t+1|t}^{22} &\equiv E_t^{\mathbb{P}}\left[(\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t+1} - E_t^{\mathbb{P}}[\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t+1}])(\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t+1} - E_t^{\mathbb{P}}[\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t+1}])'\right].\end{aligned}$$

From standard results, the evolution of $P_{t+1|t}$ is given by the recursion:

$$P_{t+1|t} = F_X\left[P_{t|t-1} - P_{t|t-1}H_X(H_X'P_{t|t-1}H_X + R)^{-1}H_X'P_{t|t-1}\right]F_X' + \ Q_X \tag{57}$$

Hence

$$\begin{aligned}& P_{t+1|t} \\ =\ & F_X\left[P_{t|t-1} - P_{t|t-1}H_X(H_X'P_{t|t-1}H_X + R)^{-1}H_X'P_{t|t-1}\right]F_X' + \ Q_X \\ =\ & F_X\left[P_{t|t-1} - \begin{pmatrix} P_{t|t-1}^{11} \\ P_{t|t-1}^{21} \end{pmatrix}(P_{t|t-1}^{11} + R)^{-1}\begin{bmatrix} P_{t|t-1}^{11} & \left(P_{t|t-1}^{21}\right)' \end{bmatrix}\right]F_X' + \ Q_X \\ =\ & F_X\left[P_{t|t-1} - \begin{bmatrix} P_{t|t-1}^{11}(P_{t|t-1}^{11} + R)^{-1}P_{t|t-1}^{11} & P_{t|t-1}^{11}(P_{t|t-1}^{11} + R)^{-1}\left(P_{t|t-1}^{21}\right)' \\ P_{t|t-1}^{21}(P_{t|t-1}^{11} + R)^{-1}P_{t|t-1}^{11} & P_{t|t-1}^{21}(P_{t|t-1}^{11} + R)^{-1}\left(P_{t|t-1}^{21}\right)' \end{bmatrix}\right]F_X' + \ Q_X\end{aligned}$$

Substitute for F_X and Q_X (see Equation (55)) yields the following expression for each block $P_{t+1|t}$:

$$P_{t+1|t}^{11} = P_{t|t-1}^{22} - P_{t|t-1}^{21}(P_{t|t-1}^{11} + R)^{-1}\left(P_{t|t-1}^{21}\right)' + H'\Sigma\Sigma'H,$$

$$\begin{aligned} P^{21}_{t+1|t} &= \left(K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}}\right)\left(P^{22}_{t|t-1} - P^{21}_{t|t-1}(P^{11}_{t|t-1}+R)^{-1}\left(P^{21}_{t|t-1}\right)'\right) + H'\Sigma_{\mathcal{E}}\Sigma' H \\ &= \left(K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}}\right)\left(P^{11}_{t+1|t} - H'\Sigma\Sigma' H\right) + H'\Sigma_{\mathcal{E}}\Sigma' H, \qquad (58) \end{aligned}$$

and

$$\begin{aligned} P^{22}_{t+1|t} &= \left(K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}}\right)\left(P^{22}_{t|t-1} - P^{21}_{t|t-1}(P^{11}_{t|t-1}+R)^{-1}\left(P^{21}_{t|t-1}\right)'\right)\left(K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}}\right)' + H'\Sigma_{\mathcal{E}}\Sigma'_{\mathcal{E}} H \\ &= \left(K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}}\right)\left(P^{11}_{t+1|t} - H'\Sigma\Sigma' H\right)\left(K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}}\right)' + H'\Sigma_{\mathcal{E}}\Sigma'_{\mathcal{E}} H. \qquad (59) \end{aligned}$$

Substitute for $P^{2}_{t+1|t}$ and $P^{22}_{t+1|t}$, define $\Phi^{\mathbb{P}}_{1\mathcal{P}} \equiv K^{\mathbb{P}}_{1\mathcal{P}} + I_{\mathcal{J}}$, and solve for $P^{11}_{t+1|t}$:

$$\begin{aligned} & P^{11}_{t+1|t} \\ = \;& P^{22}_{t|t-1} - P^{21}_{t|t-1}(P^{11}_{t|t-1}+R)^{-1}\left(P^{21}_{t|t-1}\right)' + H'\Sigma\Sigma' H \\ = \;& \Phi^{\mathbb{P}}_{1\mathcal{P}}\left(P^{11}_{t|t-1} - H'\Sigma\Sigma' H\right)\Phi^{\mathbb{P}\prime}_{1\mathcal{P}} + H'\Sigma\Sigma' H + H'\Sigma_{\mathcal{E}}\Sigma'_{\mathcal{E}} H \\ - \;& \left(\Phi^{\mathbb{P}}_{1\mathcal{P}}\left(P^{11}_{t|t-1} - H'\Sigma\Sigma' H\right) + H'\Sigma_{\mathcal{E}}\Sigma' H\right)(P^{11}_{t|t-1}+R)^{-1}\left(\Phi^{\mathbb{P}}_{1\mathcal{P}}\left(P^{11}_{t|t-1} - H'\Sigma\Sigma' H\right) + H'\Sigma_{\mathcal{E}}\Sigma' H\right)' \end{aligned}$$

Use this expression for $P^{11}_{t+1|t}$ to substitute in Equation (56):

$$\begin{aligned} E^{\mathbb{P}}_t[\mathcal{P}^o_{t+1}] &= H'_X\mu_X + H'_X F_X\left\{E^{\mathbb{P}}_{t-1}[\mathcal{X}_t] + P_{t|t-1}H_X(H'_X P_{t|t-1}H_X + R)^{-1}\left(\mathcal{P}^o_t - E^{\mathbb{P}}_{t-1}[\mathcal{P}^o_t]\right)\right\} \\ &= \begin{bmatrix} 0 & I \end{bmatrix}\begin{pmatrix} E^{\mathbb{P}}_{t-1}[\mathcal{P}_t] + P^{11}_{t|t-1}(P^{11}_{t|t-1}+R)^{-1}\left(\mathcal{P}_t - E^{\mathbb{P}}_{t-1}[\mathcal{P}^o_t]\right) \\ E^{\mathbb{P}}_{t-1}[\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t}] + P^{21}_{t|t-1}(P^{11}_{t|t-1}+R)^{-1}\left(\mathcal{P}^o_t - E^{\mathbb{P}}_{t-1}[\mathcal{P}^o_t]\right) \end{pmatrix} \\ &= E^{\mathbb{P}}_{t-1}[\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t}] + P^{21}_{t|t-1}(P^{11}_{t|t-1}+R)^{-1}\left(\mathcal{P}^o_t - E^{\mathbb{P}}_{t-1}[\mathcal{P}^o_t]\right). \end{aligned}$$

Equation (52) implies that

$$\begin{aligned} E^{\mathbb{P}}_t[\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t+1}] &= K^{\mathbb{P}}_{0\mathcal{P}} + \Phi^{\mathbb{P}}_{1\mathcal{P}}E^{\mathbb{P}}_t\left[\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t}\right], \\ E^{\mathbb{P}}_t\left[\mathcal{P}_{t+1}\right] &= E^{\mathbb{P}}_t\left[\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t}\right], \end{aligned}$$

hence

$$E^{\mathbb{P}}_{t-1}[\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t}] = K^{\mathbb{P}}_{0\mathcal{P}} + \Phi^{\mathbb{P}}_{1\mathcal{P}}E^{\mathbb{P}}_{t-1}\left[\mathcal{P}_t\right],$$

while Equation (53) implies that

$$E^{\mathbb{P}}_{t-1}\left[\mathcal{P}_t\right] = E^{\mathbb{P}}_{t-1}\left[\mathcal{P}^o_t\right],$$

thus

$$\begin{aligned} E^{\mathbb{P}}_t[\mathcal{P}^o_{t+1}] &= E^{\mathbb{P}}_{t-1}[\mathcal{E}^{\mathbb{P}}_{\mathcal{P},t}] + P^{21}_{t|t-1}(P^{11}_{t|t-1}+R)^{-1}\left(\mathcal{P}^o_t - E^{\mathbb{P}}_{t-1}[\mathcal{P}^o_t]\right) \\ &= K^{\mathbb{P}}_{0\mathcal{P}} + \Phi^{\mathbb{P}}_{1\mathcal{P}}E^{\mathbb{P}}_{t-1}\left[\mathcal{P}^o_t\right] + P^{21}_{t|t-1}(P^{11}_{t|t-1}+R)^{-1}\left(\mathcal{P}^o_t - E^{\mathbb{P}}_{t-1}[\mathcal{P}^o_t]\right). \end{aligned}$$

Use these two results to substitute in

$$E^{\mathbb{P}}_t[\mathcal{P}^o_{t+1}] = K^{\mathbb{P}}_{0\mathcal{P}} + \Phi^{\mathbb{P}}_{1\mathcal{P}}E^{\mathbb{P}}_{t-1}\left[\mathcal{P}^o_t\right] + \left(\Phi^{\mathbb{P}}_{1\mathcal{P}}\left(P^{11}_{t|t-1} - H'\Sigma\Sigma' H\right) + H'\Sigma_{\mathcal{E}}\Sigma' H\right)(P^{11}_{t|t-1}+R)^{-1}\left(\mathcal{P}^o_t - E^{\mathbb{P}}_{t-1}[\mathcal{P}^o_t]\right).$$

where we also used Equation (58). Using another well-known result from the Kalman filter:

$$\begin{aligned} \Omega_{t+1|t} &\equiv E^{\mathbb{P}}_t\left[(\mathcal{P}^o_{t+1} - E^{\mathbb{P}}_t[\mathcal{P}^o_{t+1}])(\mathcal{P}^o_{t+1} - E^{\mathbb{P}}_t[\mathcal{P}^o_{t+1}])'\right] \\ &= H'_X P_{t|t-1}H_X + R, \end{aligned}$$

we get:

$$\begin{aligned}\Omega_{t+1|t} &= H_X' P_{t+1|t} H_X + R = \begin{bmatrix} I_{\mathcal{J}} & 0 \end{bmatrix} \begin{bmatrix} P_{t+1|t}^{11} & \left(P_{t+1|t}^{21}\right)' \\ P_{t+1|t}^{21} & P_{t+1|t}^{22} \end{bmatrix} \begin{bmatrix} I_{\mathcal{J}} \\ 0 \end{bmatrix} + R \\ &= P_{t+1|t}^{11} + R.\end{aligned}$$

Define:

$$\begin{aligned}\mathcal{E}_{\mathcal{P}^o,t}^{\mathbb{P}} &\equiv E_t^{\mathbb{P}}[\mathcal{P}_{t+1}^o] \\ \Sigma_{\mathcal{P},t}^o &\equiv \Omega_{t+1|t}^{1/2} \\ \Sigma_{\mathcal{E}_{\mathcal{P}},t}^o &\equiv \left(\Phi_{1\mathcal{P}}^{\mathbb{P}}\left(P_{t+1|t}^{11} - H'\Sigma\Sigma'H\right) + H'\Sigma_{\mathcal{E}}\Sigma'H\right)\left(\Sigma_{\mathcal{P},t}^{o\prime}\right)^{-1} \\ K_{0\mathcal{P}^o}^{\mathbb{P}} &\equiv K_{0\mathcal{P}}^{\mathbb{P}} = H'K_0^{\mathbb{P}} - K_{1\mathcal{P}}^{\mathbb{P}} H_0.\end{aligned}$$

The observed portfolios have the following Conditional Mean dynamics:

$$\begin{aligned}\mathcal{P}_{t+1}^o &= \mathcal{E}_{\mathcal{P}^o,t}^{\mathbb{P}} + \Sigma_{\mathcal{P},t}^o \epsilon_{\mathcal{P}t+1}^{\mathbb{P}}, \\ \Delta\mathcal{E}_{\mathcal{P}^o,t+1}^{\mathbb{P}} &= K_{0\mathcal{P}^o}^{\mathbb{P}} + K_{1\mathcal{P}}^{\mathbb{P}}\mathcal{E}_{\mathcal{P}^o,t}^{\mathbb{P}} + \Sigma_{\mathcal{E}_{\mathcal{P}},t}^o \epsilon_{\mathcal{P}t+1}^{\mathbb{P}},\end{aligned}$$

with i.i.d. $\epsilon_{\mathcal{P}t+1}^{\mathbb{P}} \sim N(0, I_{\mathcal{J}})$. Meanwhile the variances $(\left(\Sigma_{\mathcal{P},t}^o \Sigma_{\mathcal{P},t}^{o\prime}\right)$ and $\left(\Sigma_{\mathcal{E},t}^o \Sigma_{\mathcal{E},t}^{o\prime}\right))$ are time-varying. Standard results show that $P_{t+1|t}$ converges with t, $P_{t+1|t} \to P$, under technical conditions (see e.g., Hansen and Sargent (2014)). In our case, using Proposition 13.1 in Hamilton (1994), we have that if the eigenvalues of $K_1^{\mathbb{P}}$ are all inside $(-2, 0)$, then the sequence $P_{t+1|t}^{11}$ converges to a steady-state matrix P satisfying

$$\begin{aligned}P &= \Phi_{1\mathcal{P}}^{\mathbb{P}}\left(P - H'\Sigma\Sigma'H\right)\Phi_{1\mathcal{P}}^{\mathbb{P}\prime} + H'\Sigma\Sigma'H + H'\Sigma_{\mathcal{E}}\Sigma_{\mathcal{E}}'H \\ &\quad - \left(\Phi_{1\mathcal{P}}^{\mathbb{P}}\left(P - H'\Sigma\Sigma'H\right) + H'\Sigma_{\mathcal{E}}\Sigma'H\right)(P+R)^{-1}\left(\Phi_{1\mathcal{P}}^{\mathbb{P}}\left(P - H'\Sigma\Sigma'H\right) + H'\Sigma_{\mathcal{E}}\Sigma'H\right)'. \qquad (60)\end{aligned}$$

Equation (60) is a generalized algebraic Ricatti equation. A numerical solution is readily available in most scientific software, including MATLAB, if a solution exists (see Arnold and Laub 1984). Substitute for $P_{t+1|t}^{11}$ using its steady-state value P given in Equation (60) to obtain the Conditional Mean of observed portfolios but with invariant parameters:

$$\begin{aligned}\mathcal{P}_{t+1}^o &= \mathcal{E}_{\mathcal{P}^o,t}^{\mathbb{P}} + \Sigma_{\mathcal{P}}^o \epsilon_{\mathcal{P}t+1}^{\mathbb{P}}, \\ \Delta\mathcal{E}_{\mathcal{P}^o,t+1}^{\mathbb{P}} &= K_{0\mathcal{P}^o}^{\mathbb{P}} + K_{1\mathcal{P}^o}^{\mathbb{P}}\mathcal{E}_{\mathcal{P}^o,t}^{\mathbb{P}} + \Sigma_{\mathcal{E}_{\mathcal{P}}}^o \epsilon_{\mathcal{P}t+1}^{\mathbb{P}} \\ K_{0\mathcal{P}^o}^{\mathbb{P}} &\equiv K_{0\mathcal{P}}^{\mathbb{P}} = H'K_0^{\mathbb{P}} - K_{1\mathcal{P}}^{\mathbb{P}} H_0,\end{aligned}$$

where

$$\begin{aligned}K_{1\mathcal{P}^o}^{\mathbb{P}} &\equiv K_{1\mathcal{P}}^{\mathbb{P}} = H'(K_1^{\mathbb{P}} + I_{\mathcal{N}})H^{\perp} - I_{\mathcal{J}} \\ \Sigma_{\mathcal{P}}^o \Sigma_{\mathcal{P}}^{o\prime} &\equiv P + R\end{aligned}$$

$$\begin{aligned}\Sigma_{\mathcal{E}_{\mathcal{P}}}^o &\equiv \left(\Phi_{1\mathcal{P}}^{\mathbb{P}} P - \Phi_{1\mathcal{P}}^{\mathbb{P}} H'\Sigma\Sigma'H + H'\Sigma_{\mathcal{E}}\Sigma'H\right)\left(\Sigma_{\mathcal{P}}^{o\prime}\right)^{-1} \\ &= \left(\Phi_{1\mathcal{P}}^{\mathbb{P}} P - H'(K_1^{\mathbb{P}} + I_{\mathcal{N}})H^{\perp}H'\Sigma\Sigma'H + H'\Sigma_{\mathcal{E}}\Sigma'H\right)\left(\Sigma_{\mathcal{P}}^{o\prime}\right)^{-1}\end{aligned}$$

$$
\begin{aligned}
&= \left(\Phi^{\mathbb{P}}_{1\mathcal{P}}P - H'\left((K^{\mathbb{P}}_1 + I_{\mathcal{N}})\Sigma\Sigma' + \Sigma_{\mathcal{E}}\Sigma'\right)H\right)\left(\Sigma^{o'}_{\mathcal{P}}\right)^{-1} \\
&= \left(\Phi^{\mathbb{P}}_{1\mathcal{P}}P - H'\left((K^{\mathbb{P}}_1 + I_{\mathcal{N}}) + \Sigma_{\mathcal{E}}\Sigma^{-1}\right)\Sigma\Sigma'H\right)\left(\Sigma^{o'}_{\mathcal{P}}\right)^{-1} \\
&= \left(\Phi^{\mathbb{P}}_{1\mathcal{P}}P - H'\left((K^{\mathbb{P}}_1 + I_{\mathcal{N}}) + \Sigma_{\mathcal{E}}\Sigma^{-1}\right)\Sigma\Sigma'H\right)\left(\Sigma^{o'}_{\mathcal{P}}\right)^{-1} \\
&= \left(\Phi^{\mathbb{P}}_{1\mathcal{P}}P - H'\theta\Sigma\Sigma'H\right)\left(\Sigma^{o'}_{\mathcal{P}}\right)^{-1}
\end{aligned}
$$

which is the result stated in Theorem 1 but where we drop the superscript ^{o} from the parametrization. We use the superscript in this proof to differentiate parameters for the dynamics of the unobserved $\mathcal{P}_t$ from the parameters of the dynamics of $\mathcal{P}^o_t$. This distinction is not required in the text since we never introduce the parameters for $\mathcal{P}_t$ there. However, note that we always differentiate between $\mathcal{P}_t$ and $\mathcal{P}^o_t$ themselves in the text.

A.4 Calibration

From the recursion in Equation 45:

$$
\mathcal{A}_n = \left(\sum_{j=0}^{n-1}\mathcal{B}'_j\right)K^{\mathbb{Q}}_0 + \frac{1}{2}\left(\sum_{j=0}^{n-1}\mathcal{B}'_j\Sigma\Sigma'\mathcal{B}_j\right) - n\rho_0. \tag{61}
$$

Consider a set of maturities, $n_1, \cdots, n_{\mathcal{J}}$ and stack coefficients into vectors,

$$
\mathcal{A} = -\begin{bmatrix} \mathcal{A}_{n_1}/n_1 \\ \mathcal{A}_{n_2}/n_2 \\ \vdots \\ \mathcal{A}_{n_{\mathcal{J}}}/n_{\mathcal{J}} \end{bmatrix} \quad \mathcal{B} \equiv -\begin{bmatrix} \mathcal{B}'_{n_1}/n_1 \\ \mathcal{B}'_{n_2}/n_2 \\ \vdots \\ \mathcal{B}'_{n_{\mathcal{J}}}/n_{\mathcal{J}} \end{bmatrix}.
$$

Using equation (61), we have

$$
\mathcal{A} = \rho_0 + \mathcal{D}K^{\mathbb{Q}}_0 + \mathcal{C}
$$

with

$$
\mathcal{D} = -\begin{bmatrix} \left(\sum_{j=0}^{n_1-1}\mathcal{B}'_j\right)/n_1 \\ \left(\sum_{j=0}^{n_2-1}\mathcal{B}'_j\right)/n_2 \\ \vdots \\ \left(\sum_{j=0}^{n_J-1}\mathcal{B}'_j\right)/n_{\mathcal{J}} \end{bmatrix} \quad \mathcal{C} = -\frac{1}{2}\begin{bmatrix} \left(\sum_{j=0}^{n_1-1}\mathcal{B}'_j\Sigma\Sigma'\mathcal{B}_j\right)/n_1 \\ \left(\sum_{j=0}^{n_2-1}\mathcal{B}'_j\Sigma\Sigma'\mathcal{B}_j\right)/n_2 \\ \vdots \\ \left(\sum_{j=0}^{n_J-1}\mathcal{B}'_j\Sigma\Sigma'\mathcal{B}_j\right)/n_{\mathcal{J}} \end{bmatrix}.
$$

Therefore, the unconditional expectation is given by

$$
E\left[\mathcal{P}^o_t\right] = H_0 = W\left(\rho_0 + \mathcal{D}K^{\mathbb{Q}}_0 + \mathcal{C}\right),
$$

since $K^{\mathbb{P}}_0 = 0$ and $H_0 = W\mathcal{A}$ where W is a fixed known and invertible $\mathcal{J}\times\mathcal{J}$ matrix. We want to match the unconditional expectation of a subset $W_1\mathcal{P}^o_t$ of the portfolios $\mathcal{P}^o_t$, where W_1 is a $\mathcal{N}\times\mathcal{J}$ matrix:

$$
E\left[W_1\mathcal{P}^o_t\right] = W_1H_0 = W_1W\left(\rho_0 + \mathcal{D}K^{\mathbb{Q}}_0 + \mathcal{C}\right) = \rho_0W_1W + W_1W\mathcal{D}K^{\mathbb{Q}}_0 + W_1W\mathcal{C},
$$

which leads to

$$
K^{\mathbb{Q}}_0 = \left(W_1W\mathcal{D}\right)^{-1}\left[E\left[W_1\mathcal{P}^o_t\right] - \rho_0W_1W - W_1W\mathcal{C}\right].
$$

Figure 1: Model selection criteria favors a parsimonious long-lag structure.
Hannan-Quinn and BIC model selection criteria for $VAR(p)$ models with $p = 1, \ldots, 12$ in Panel (A) and Panel (B), respectively. The corresponding criteria for the $VARMA(1,1)$ is also reported (dashed) for comparison. Estimation by maximum likelihood based on a sample from January 1964 until December 2007.

(A) Hannan-Quinn Criteria **(B)** BIC Criteria

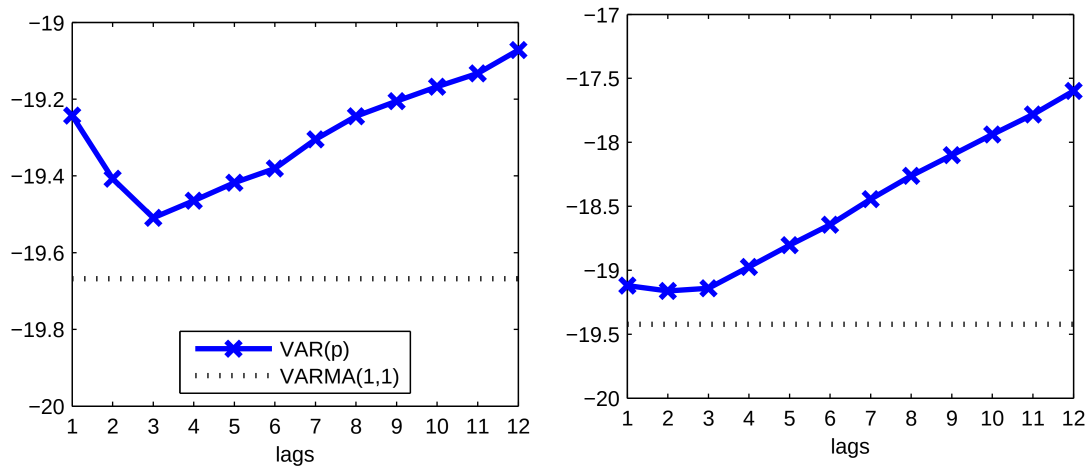


Figure 2: Fitting the Cochrane-Piazzesi Forecast in 5-factor Models
Regressions of the returns forecasting factor on the model-implied bond risk premia

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

where $\hat{xr}_{t,h}^{(n)} = \hat{b}_{n,h}\mathcal{R}_{t,h}$ is estimated from Equation 3. We report the ratio of the R^2s using the CM_5^0 model relative to the R^2s from the CM_5^5 model. Estimation based on a sample from January 1964 until December 2007. Returns computed from the GSW data set.

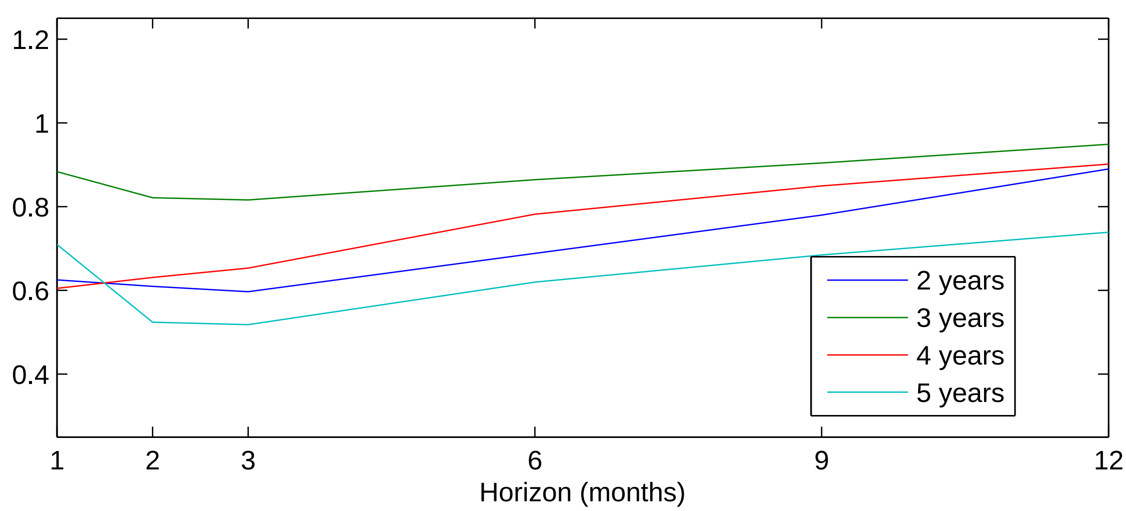
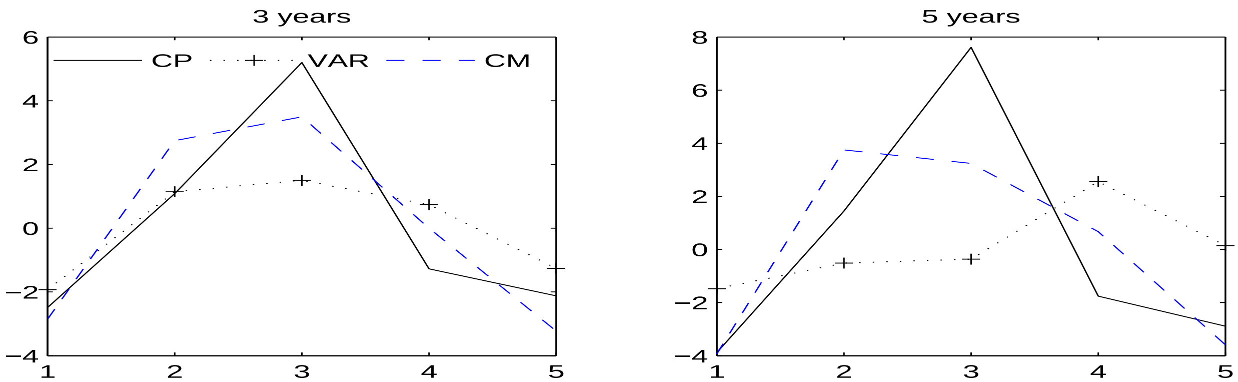


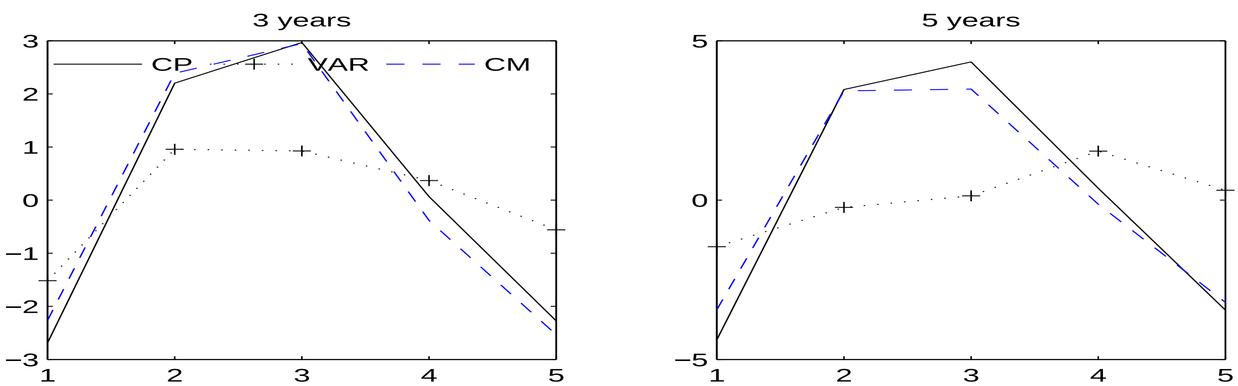
Figure 3: Predictability Coefficients in 5-Factor Models

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

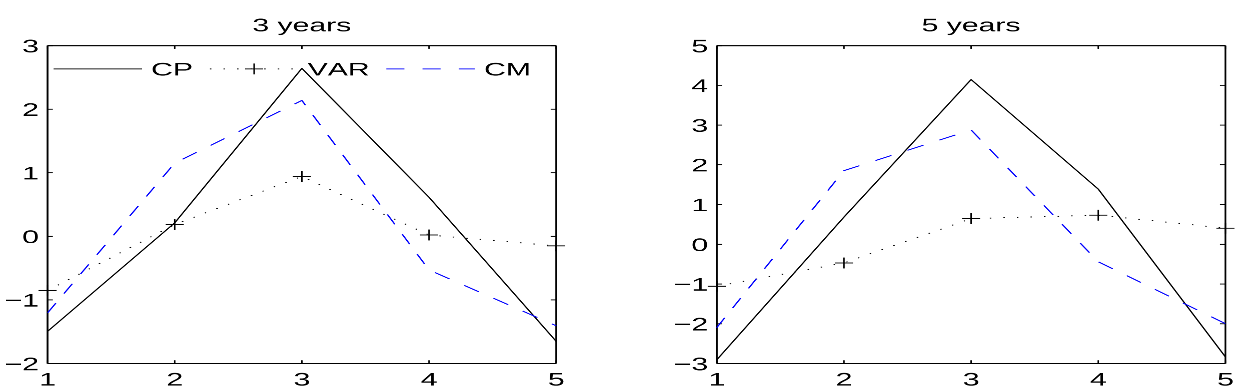
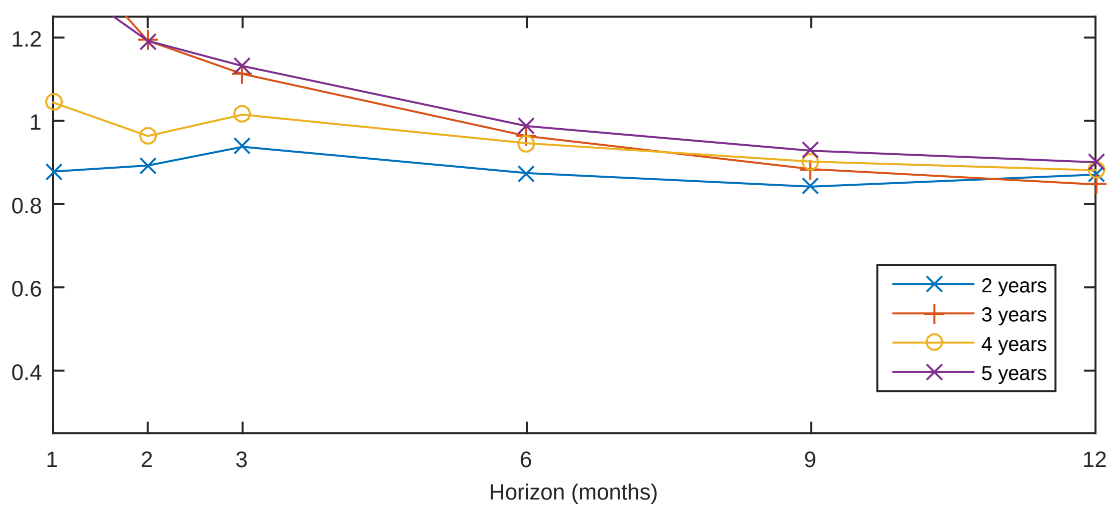


Figure 4: Fitting the Cochrane-Piazzesi Forecast in 3-factor Models

Regressions of the returns forecasting factors on the model-implied bond risk premia

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

where $\hat{xr}^{(n)}_{t,h} = \hat{b}_{n,h}\mathcal{R}_{t,h}$ is estimated from Equation 3. We report the ratio of the R^2s relative to the R^2s obtained using risk premium from the CM_5^5-$KF0$ model. Panel (A) reports results for the CM_3^3-$KF0$ model. Panel (B) reports results for the CM_3^3-$KF1$ model. Estimation based on a sample from January 1964 until December 2007. Returns computed from the GSW data set.



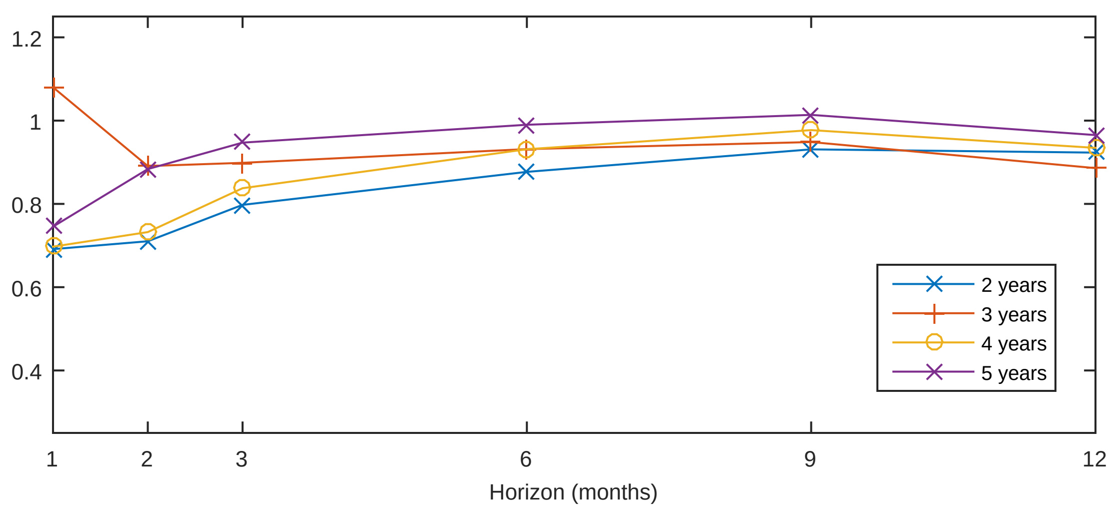


Figure 5: Term Premium

Term premium in 5-year yield implied by the CM_3^3-$KF1$, CM_3^2-$KF1$ and CM_3^1-$KF1$ models. Estimation based on a sample from January 1964 until December 2007.

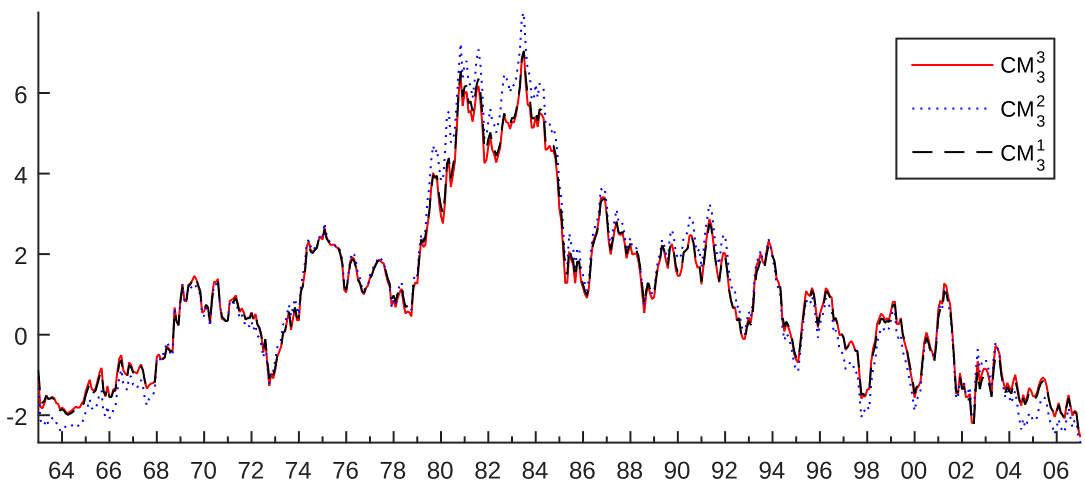


Figure 6: Yield Decompositions and Term Premium

Decomposition of the 5-year yield implied by the CM_3^0-$KF0$ and CM_3^3-$KF1$ models. Estimation based on a sample from January 1964 until December 2007. The level of the expectation term is fixed to the 5-year yield sample average at estimation.

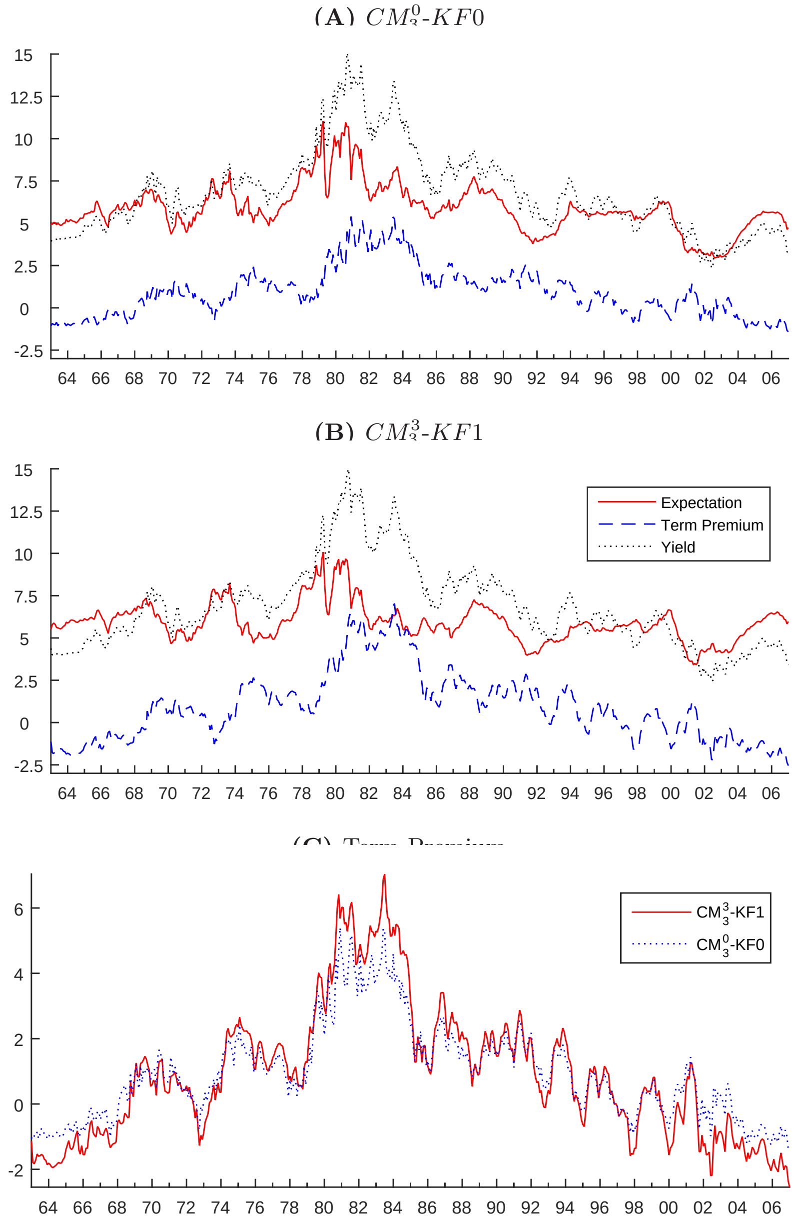


Table 1: **Bond Excess Returns – OLS Predictability Regressions**

Predictability of excess returns $xr_{t,h}^{(n)}$ on bonds with $n = 2, 3, 4, 5, 7, 10$ years to maturity for holding horizons between t and $t + h$ with $h = 1, 2, 3, 6, 9, 12$ months, and 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,

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

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) CP Regressions with Distributed Lags

	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 $h = 1, 2, 3, 6, 9, 12$ 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

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

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

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

estimated across horizons of 1, 2, 3, 6, 9, and 12 months. Returns computed monthly from the GSW data set between Jan. 1963 and Dec. 2003.

	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	0.3	0.1	0.02	0.003

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

Panel (A) reports coefficient estimates in regressions of each element of $\hat{w}_t$ on its own lag $\hat{w}_{t-h}$, where $\hat{w}_t$ is the estimated residual from a Markovian model estimated for the yield vector Y_t, and where we vary $h = 1, 3, 5$. Panel (B) reports coefficient estimates in univariate regressions of the element of $\hat{w}_t$ on every element of the first lag separately. Panels (C) and (D) report coefficient estimates $\hat{\gamma}_\epsilon$ and $\hat{\alpha}_\epsilon$ from a two-step estimation of the single-factor regression $\hat{w}_t = \alpha_w \gamma_w' \ \hat{w}_{t-1} + \eta_t$, 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

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

Panel (B) Cross-correlations

	$w_t^{(1)}$	$w_t^{(2)}$	$w_t^{(3)}$	$w_t^{(4)}$	$w_t^{(5)}$
$w_{t-1}^{(1)}$	0.15	0.14	0.09	0.06	0.06
	(1.77)	(2.00)	(1.50)	(1.15)	(1.18)
$w_{t-1}^{(2)}$	0.18	0.17	0.11	0.08	0.08
	(2.06)	(2.25)	(1.72)	(1.36)	(1.44)
$w_{t-1}^{(3)}$	0.20	0.19	0.12	0.10	0.09
	(2.44)	(2.72)	(2.06)	(1.65)	(1.83)
$w_{t-1}^{(4)}$	0.20	0.19	0.12	0.09	0.09
	(2.19)	(2.43)	(1.83)	(1.47)	(1.61)
$w_{t-1}^{(5)}$	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

	$w_{t-1}^{(1)}$	$w_{t-1}^{(2)}$	$w_{t-1}^{(3)}$	$w_{t-1}^{(4)}$	$w_{Y,t-1}^{(5)}$	p-val
$\hat{\gamma}_\epsilon^{(n)}$	-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

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

Table 5: **Out-of-Sample forecast RMSEs favors a parsimonious long-lag model.**
Ratios of out-of-sample forecast RMSEs relative to the $VAR(1)$ model. Forecasts of future yields with annual maturity from one to five years (across rows) at horizons of 1, 3, 6, 9 and 12 months (across columns). Starting in 1985, each model is estimated with data until the end of the year to produce yield forecasts up to twelve months ahead. Each year, all models are re-estimated with one year of additional data. Panels (A)-(D): results for the $VARMA(1,1)$, $VAR(1)$, $VAR(2)$ and $VAR(12)$ models, respectively. Estimation based on maximum likelihood based on data between Jan. 1963 and Dec. 2007.

Panel (A) $VARMA(1,1)$ vs VAR(1)

	Forecast Horizon (months)			
	3	**6**	**9**	**12**
1	1.17	1.16	1.11	1.06
2	1.07	1.10	1.05	1.00
3	0.98	1.02	0.99	0.94
4	0.93	0.98	0.95	0.89
5	0.90	0.95	0.92	0.87

Panel (B) $VAR(2)$ vs VAR(1)

	Forecast Horizon (months)			
	3	**6**	**9**	**12**
1	1.42	1.40	1.33	1.26
2	1.39	1.38	1.31	1.25
3	1.24	1.31	1.27	1.23
4	1.17	1.28	1.25	1.21
5	1.08	1.24	1.23	1.19

Panel (C) $VAR(3)$ vs VAR(1)

	3	**6**	**9**	**12**
1	1.64	1.51	1.43	1.34
2	1.62	1.49	1.41	1.34
3	1.47	1.39	1.35	1.32
4	1.49	1.37	1.33	1.30
5	1.45	1.32	1.29	1.27

Panel (D) $VAR(12)$ vs VAR(1)

	3	**6**	**9**	**12**
1	2.40	1.69	1.30	1.16
2	2.46	1.65	1.29	1.15
3	2.44	1.57	1.28	1.15
4	2.62	1.62	1.32	1.16
5	2.71	1.65	1.31	1.15

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 (28) 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, the gain in likelihood relative to the most restricted CM_3^0-$KF0$, and the gain in BIC model selection criteria relative to the CM_3^0-$KF0$. 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-$KF0$ model for bond maturities 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.

	$KF0$		$KF1$				$KF2$			
	CM_3^0	CM_3^3	CM_3^0	CM_3^1	CM_3^2	CM_3^3	CM_3^0	CM_3^1	CM_3^2	CM_3^3
	Panel (A) Changes in parameters, likelihood and selection criteria									
	19	25	19	22	24	25	23	26	28	29
	0	20	100	121	126	128	314	329	332	334
	0	3	201	223	221	219	602	615	608	606
	Panel (B) Variance ratios									
2	1.23	0.87	1.04	0.87	1.07	1.07	0.84	0.83	0.85	0.84
3	1.30	0.92	0.99	0.80	1.05	1.04	0.78	0.75	0.78	0.78
4	1.30	0.92	0.97	0.77	1.05	1.02	0.76	0.71	0.76	0.75
5	1.42	0.93	1.02	0.79	1.12	1.08	0.80	0.74	0.79	0.79
	Panel (C) R^2 ratios									
2	0.71	1.00	0.73	0.76	0.77	0.77	0.70	0.66	0.66	0.67
3	0.65	1.03	0.77	0.82	0.82	0.81	0.69	0.67	0.66	0.68
4	0.66	1.05	0.86	0.92	0.89	0.88	0.73	0.72	0.71	0.74
5	0.60	1.05	0.84	0.86	0.88	0.86	0.61	0.59	0.59	0.63

Table 8: **The improvements in out-of-sample forecast accuracy relative to a 5-factor model**
Ratios of out-of-sample forecasts RMSEs from $CM_{\mathcal{N}}^{k}$-KF models with $\mathcal{N}=3$ relative to the CM_5^5-$KF0$ model. A value less than one indicates more accurate forecasts. Ratios for $KF1$ and $KF2$ models in the first and second columns, respectively. Estimation is repeated every month with data since January 1964 and with an expanding end-of-sample date, starting in December 1985 and until December 2015. Data from CRSP.

Panel (A) CM_3^1-$KF1$

	Forecast Horizon (months)				
	1	3	6	9	12
1	0.994	0.949	0.945	0.961	0.974
2	1.045	0.976	0.952	0.955	0.961
3	1.058	0.988	0.957	0.951	0.954
4	1.015	0.974	0.948	0.944	0.947
5	1.017	0.986	0.965	0.955	0.952

Panel (B) CM_3^1-$KF2$

	Forecast Horizon (months)				
	1	3	6	9	12
1	1.045	0.991	0.967	0.975	0.980
2	1.126	1.003	0.964	0.964	0.967
3	1.096	1.002	0.962	0.958	0.960
4	1.050	0.982	0.951	0.951	0.956
5	1.051	0.998	0.971	0.966	0.965

Panel (C) CM_3^2-$KF1$

	1	3	6	9	12
1	1.010	0.957	0.953	0.970	0.983
2	1.049	0.981	0.960	0.964	0.972
3	1.059	0.993	0.952	0.962	0.965
4	1.011	0.977	0.956	0.954	0.959
5	1.012	0.989	0.973	0.965	0.964

Panel (D) CM_3^2-$KF2$

	1	3	6	9	12
1	1.029	1.015	0.952	1.022	1.025
2	1.087	1.020	0.952	1.012	1.015
3	1.062	1.019	0.952	1.011	1.013
4	1.015	0.996	0.952	1.004	1.011
5	1.018	1.007	1.005	1.015	1.018

Panel (E) CM_3^3-$KF1$

	1	3	6	9	12
1	1.009	0.958	0.956	0.973	0.986
2	1.050	0.981	0.962	0.967	0.974
3	1.060	0.993	0.967	0.964	0.968
4	1.017	0.979	0.959	0.957	0.962
5	1.013	0.990	0.976	0.969	0.967

Panel (F) CM_3^3-$KF2$

	1	3	6	9	12
1	1.012	0.992	0.985	0.998	1.004
2	1.086	1.004	0.983	0.987	0.992
3	1.067	1.006	0.985	0.984	0.987
4	1.019	0.988	0.976	0.978	0.984
5	1.022	1.002	0.995	0.992	0.992

Table 9: **Conditional Mean Models with Five Factors**

Results from $CM_{\mathcal{N}}^{k}$-KF models with $\mathcal{N} = 5$. The first line shows the label of each model. Panel (A) reports the number of parameters for each model, the gain in likelihood and the gain in BIC model selection criteria relative to the most restricted model CM_5^0-$KF0$. 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-$KF0$ model 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_5^0			CM_5^5
	$KF0$	$KF1$	$KF2$	$KF0$
Panel (A) Changes in parameters, likelihood and selection criteria				
	45	46	50	70
	0	38	60	98
	0	70	89	40
Panel (B) Variance ratios				
2	0.84	0.69	0.70	1.00
3	0.82	0.67	0.73	1.00
4	0.84	0.66	0.73	1.00
5	0.89	0.67	0.69	1.00
Panel (C) R^2 ratios				
2	0.75	0.81	1.05	1.00
3	0.75	0.81	1.00	1.00
4	0.77	0.85	1.01	1.00
5	0.65	0.72	0.99	1.00