Gaussian Term Structure Models and Bond Risk Premia

Bruno Feunou  Jean-Sébastien Fontaine
Bank of Canada

July 2013

Abstract

Cochrane and Piazzesi (2005) show that lagged forward rates add to current forward rates, improving the predictability of annual bond returns, and that a Markovian model applied to forward rates at the monthly frequency cannot generate the pattern of predictability in annual returns. These results stand as a challenge to dynamic modern term structure models (DTSM). We develop the family of Conditional Mean DTSM where the time-series dynamics of yield factors depend on current yields and on their history. Empirically, we find that both current and past yields generate substantial bond risk premium variations: the model-implied premiums are close to the return-forecasting factors obtained from direct predictability regressions. In addition, the population coefficients from our monthly model are close to the tent-shaped coefficient from CP regressions across different returns horizons. Finally, we find that since 2007 the risk premium on the 10-year yield is lower, and the expectation component is higher, by 0.5% than what a standard Markovian model would suggest.

Keywords: Term Structure Models
JEL Classification: E43, E47, G12

We thank Antonio Diez, Greg Duffee, Anh Le and Dora Xia and participants to the 4th Bank of Canada Fixed Income conference for comments and suggestions. Email: bfeunou@bankofcanada.ca or jsfontaine@bankofcanada.ca.
Introduction

Decomposing the yields to maturity of Treasury securities into a pure expectation component and the compensation for risk offered to bondholders is the *raison d’être* of modern Dynamic Term Structure Models (DTSMs). In the absence of a risk premium, we can interpret the yield curve as the bond investors’ expectation of the short rate at future dates. But returns from holding long-dated bonds, in excess of returns from holding a short-dated bond to maturity, exhibits predictable variations. The evidence “stands as challenges or ‘stylized facts’ to be explained by candidate models” (Fama, 1984b) but, using the Campbell and Shiller (1991) regressions as a benchmark, Dai and Singleton (2002) show that Gaussian DTSMs (GDTSM) meet this challenge.\footnote{See also Shiller (1979); Startz (1982); Fama (1984a,b); Fama and Bliss (1987); Campbell and Shiller (1991) for earlier evidence against the expectation hypothesis. Dai and Singleton (2002) show that GDTSMs (i) can reproduce the pattern of coefficients from regressions of yield changes on the yield curve slope and that (ii) coefficients from regressions of ‘risk-adjusted’ yield changes are consistent with the expectation hypothesis.}

Cochrane and Piazzesi (2005) (CP) show that the predictive content of lagged lagged forward rates add to the predictability of bond returns.\footnote{See Cochrane and Piazzesi (2005), Section III.B and III.C, pp. 151-153.} CP also show that a Markovian model on forward rates at the monthly frequency cannot produce the pattern of predictability at longer horizons. The first result presents a challenge to DTSM because the defining feature of a VAR(1) is that current yields contain all the information about expected future yields. In CP’s 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”. The second results present yet an additional but distinct challenge to the ability of standard DTSMs to accurately aggregate returns forecasts across longer horizons. This paper develops a parsimonious non-Markovian extension of the standard Gaussian DTSM capturing the role of past yields in explaining bond risk premium. We also show that our model help bridge the aggregation gap between monthly forward rates and returns at longer horizons. The resulting yield decompositions differ from those obtained from Markovian specifications, with important implications for investors.

The first part of the paper establishes a set of stylized facts to motivate our model specification. Our primary interest is to establish the pattern of bond returns predictability, emphasizing the role of lagged forward rates. Unsurprisingly, we find
that a combination of forward rates summarize the predictive content of the term structure across horizons from one month to one year and for bond maturities up to ten years. More importantly, we ask whether lags of the return-forecasting factor add to the predictability of bond returns across horizons? The answer is a resounding yes. In fact, the relative contribution of lagged forward rates increases as we shorten the horizon and peaks at the monthly horizon in direct contradiction with Markovian models. In addition, we find that for every horizon, the coefficients on lagged forward rates exhibit a decaying pattern. This is consistent with the presence of a moving-average component (as shown by CP for annual returns). We also note that the horizon-specific returns-forecasting factors exhibit a tight factor structure connecting the bond risk premiums across different investment horizons. Indeed, we show that simply adding a moving-average component to the monthly VAR(1) in CP captures the predictability evidence for annual returns.

The second part of the paper introduces a Gaussian DTSM (GDTSM) to match the stylized facts. First, we clarify the fact that if yields are linear in the risk factors $Z_t$ with a stationary gaussian distribution, then no-arbitrage requires that $Z_t$ is Markovian under the risk-neutral 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. Second, it follows that the information content of lagged forward rates arises from the historical dynamics. Hence, we develop the class of Conditional Mean GDTSMs where the conditional expectation $\mathbb{E}_{Z,t} = \mathbb{E}_t[Z_{t+1}]$ is Markovian under the historical probability measure $\mathbb{P}$. Representing the evolution of yield factors in terms of their conditional mean is natural in the context of term structure. Fiorentini and Sentana (1998) introduced the general conditional mean representation in the context of time-series models. Third, we show that Conditional Mean specification under $\mathbb{P}$ is equivalent to a specification of the prices of risk that is an affine functions of $Z_t$ and $\mathbb{E}_{Z,t-1}$ (or of $\mathbb{E}_{Z,t-1}$ only). In other words, one can also see our approach as a generalization of the standard prices of risk. Importantly, this generalization implies that bond risk premiums depend on current yields as well as on the history of yields. In addition, the coefficients on lagged yields exhibiting a decaying pattern as found in the data.

The second part of the paper shows that a conditional mean GDTSM captures much of the predictability pattern of bond returns that has been documented. Adapting – and nesting – the canonical form in (Joslin, Singleton, and Zhu, 2011), we develop a maximally flexible canonical form that can be estimated easily. The results
can be summarized as follow. First, conditioning on the information content of past yields increases the variability of the model-implied risk premium. For a 2-year bond, and for investment horizons between one and twelve months, a Markovian models misses between 45% and 70% of the variability. Second, this added variability is almost entirely translated into higher accuracy in explaining measures of the risk premium based on direct regressions. Finally, the conditional mean GDTSM provides a close match to the forward rate coefficients estimated from the CP regressions while the Markovian model provides a poorer fit and does not reproduce the well-known tent-shape present in the data. This extends the results from Dai and Singleton (2002), showing that gaussian DTSMs captures the predictability evidence in Campbell and Shiller (1991). We show that the risk factors underlying the cross-section of yields have significant non-Markovian dynamics.

Turning to the decomposition of the 10-year yield. We find that the expectation components from a Conditional Mean model and from the Markovian VAR(1) models were close to each other since the recession of 2001 but diverged abruptly during the course of the second half of 2007. Since then, the 10-year expectation of future short rates from the conditional model are 0.5% higher, and that the risk premium is lower by 0.5%, than the corresponding figures from the Markovian model. This may be due to unspanned flight-to-liquidity effects (Fontaine and Garcia, 2012), or because of unconventional monetary policy in the US or elsewhere.

CP suggest that forward rates are Markovian but with small measurement errors that are poorly correlated over time. Then, it follows from the standard Kalman “that the best guess of the true state is a geometrically weighted moving average.” (See Cochrane and Piazzesi (2005), p.154.) This is consistent with recent findings in Duffee (2011) where a substantial share of the information contained in the filtered state is not spanned by current yields but it is rather obtained from the history of yields via the Kalman filter. In fact, we show that the combination of Markovian model with the Kalman filter implies that the dynamics of filtered risk factors belongs to the family of Conditional Mean GDTSMs. We then have two distinct mechanisms to generate non-Markovian dynamics. First, the true underlying risk factors are non-Markovian. Second the true risk factors are Markovian but partially hidden by measurement errors. These two channels are not mutually exclusive. In addition, we cannot be separately identified them without additional assumptions. However, we show that the presence of measurement errors on its own cannot generate the pattern of predictability that has been documented.

The class of GDTSMs includes the large family of VAR or VARMA models.
Nonetheless, as pointed out by CP, most analysts select a VAR(1) often on the ground that current yields should span all the information relevant for the evolution of future yields. This presumption is not required by the absence of arbitrage and contrasts with the predictability evidence. Ang and Piazzesi (2003) incorporate lags of macro variable within a macro-finance models (See also Ang, Dong, and Piazzesi 2007 and Jardet, Monfort, and Pegoraro 2008). Monfort and Pegoraro (2007) consider yield factor models with $0 < p < \infty$ lags and with regime-switching coefficients. We circumvent the need for regimes (non-linearities), we treat the case $p = \infty$ parsimoniously, and we show the importance of conditioning on the history of yields to reveal the risk premiums. More recently, Feunou and Fontaine (2012) show the importance of including an MA term in a macro-finance term structure model to forecast inflation (see also the discussion in Kim (2007)). Joslin, Le, and Singleton (2013) consider lags of macro variables and consider the effect of spanning restrictions on estimated monetary policy rules. Yield factors are also non-markovian in Joslin, Priebsch, and Singleton (2010) but the combination of current yield factors and macro variables is markovian. We differ in that we introduce a non-markovian model based on yield factors only.

Froot (1989) points out that the usual tests of the expectation hypothesis rely on the maintained assumption that investors’ expectations are rational. Using surveys of interest-rate expectations, Froot (1989) argues that expectational errors play a significant role for long-dated bonds. Then, lagged forward rates can play an important role in predicting the risk premiums. Piazzesi and Schneider (2009) provide additional evidence that subjective expected excess returns are less volatile and less cyclical. Recently, Cieslak and Povala (2013) also recognizes the significance of lagged information and explore the role of informational frictions related to agents’ perception of the policy rule. Johannes, Lochstoer, and Mou (2011) formally study the asset pricing implication of Bayesian learning about model the fundamental’s dynamics. Our focus is different. We note that the predictive content of lagged forward rates is a challenge to standard GDTSMs where yield factors are Markovian and argue that a Conditional Mean model reconciles a risk-based explanation with the data. Compensation for risk and informational frictions (or learning) may both be consistent with non-markovian dynamics in yields. We leave this fundamental question for future research.

---

3 Another common argument is that any dynamic with longer lags can be re-cast within an extended state representation with only one lag. See, e.g., Equation 7-8 in Ang and Piazzesi (2003) or footnote 7 in Joslin, Singleton, and Zhu (2011). This argument is valid on for analytical purposes but not for empirical purposes.
1 Evidence in Bond Risk Premia

This section extends CP’s results and establishes a set of stylized facts to motivate the subsequent development of a non-markovian model. To summarize, one combination of forward rates subsumes the predictability of bond returns across horizons between one and twelve months. Second, in each case we find that lags of the forward-based factor improve bond returns forecasts. Third, simply adding a moving-average component to the standard VAR(1) dynamics is consistent with the patterns of predictability. Finally, the horizon-specific returns-forecasting factors based on current and past forwards are highly correlated and exhibits a tight factor structure.

1.1 Cochrane-Piazzesi Regressions

CP document that a linear combination of forward rates forecasts annual excess returns from holding Treasury bonds with different maturities. They focus on annual returns and bonds with 2, 3, 4 and 5 years to maturity. Our main objective in this section is to document the pattern of returns predictability as monthly returns are aggregated over longer holding periods. For this purpose, we compute excess returns using the GSW data set. However, we consider the same predictors than CP (annual forward rates with 1, 2, 3, 4 and 5 years to maturity) and the same sample period, from 1963 until 2003. We find similar results in longer samples.

We first estimate the unrestricted predictability regressions on current forward rates,

\[ x_{t,t+h}^{(n)} = \beta_{n,h}' f_t + u_{t,h}^{(n)}, \quad (1) \]

where, following CP’s notation, \( \beta_{n,h} \) is a vector of coefficients, \( f_t \) stacks a constant with the forward rates and \( x_{t,t+h}^{(n)} \) is the excess returns from holding a bond with maturity \( n \) over an \( h \)-months horizon. Table 1 displays the results for bonds with maturities up to 10 years and across horizons from one month to one year. Panel (A) shows that the degree of predictability increases with the horizons – from around 3% at the 1-month horizon, to around 10%, 20%, 30% and 35% at the 3, 6, 9, and 12-month horizons respectively. To save space, we do not report the numerous forward rate coefficient estimates. However, the estimates exhibit a similar tent-shaped patterns at each horizon, suggesting that a single combination of forward rates summarizes the predictability across maturities. We verify that a factor structure holds via the following regressions:

\[ x_{t,t+h}^{(n)} = b_{n,h} \gamma_{h}' f_t + u_{t,t+h}^{(n)}; \quad (2) \]
where $b_{n,h}$ is a scalar and $\gamma_h$ is an horizon-specific vector of coefficients. Panel (B) of Table 1 displays $R^2$s from the restricted regressions. The results are clear: the single-factor restriction leads to essentially no loss of predictability. For any given horizon, the same linear combination of forward rates forecasts the excess returns across different bond maturities.

One natural question is whether we can collapse the horizon-specific forward rate factors estimated so far into a single factor. Indeed, the horizon-specific factors are highly correlated with each other and the first principal component extracted from the panel of factors explains 93.7% of total variations and the second component explains most of the remaining. Hence the evidence suggests one or two factors from forward rates predict bond returns across different maturities and different horizons.

### 1.2 Recursive Cochrane-Piazzesi Regressions

CP show that lags of the forward-based factor increases returns predictability significantly. Moreover, the coefficients on additional factor exhibit a decaying pattern. For instance, repeating the regression in CP of excess returns on lagged values of their returns-forecasting factors,

$$\bar{x}r_{t+12} = \gamma'(\alpha_0f_t + \cdots + \alpha_8f_{t-8}),$$

where we allow for 8 lags, we obtain the following estimates of the weights: 0.29, 0.25, 0.16, 0.14, 0.05, 0.07, 0.03 and 0.002. (Compare with their Table 5B.) One simple representation of this decaying pattern is given by:

$$\bar{x}r_{t,h}^{(n)} = (1 - \alpha_{n,h})\gamma'_{n,h}f_t + \alpha_{n,h}\mathcal{R}_{t-1,h}^{(n)} + u_{t,h}^{(n)},$$

where the coefficients $\alpha_{n,h}$ is a scalar and where $\mathcal{R}_{t-1,h}^{(n)}$ summarizes the information from past forward rates is given by the following recursions:

$$\mathcal{R}_{t,h}^{(n)} = \alpha_{n,h}\mathcal{R}_{t-1,h}^{(n)} + (1 - \alpha_{n,h})\gamma'_{n,h}f_t,$$

implying the following for bond risk premia:

$$\bar{x}r_{t,h}^{(n)} = (1 - \alpha_{n,h})\sum_{j=0}^{\infty} \alpha_{n,h}^j (\gamma'_{n,h}f_{t-j}) + u_{t,h}^{(n)},$$

---

4We proceed via iterated estimation over $\gamma$ and $\alpha$ as in CP.
Equations 5 captures, in a simple reduced-form way, the observation in CP that lagged forward rates matter but with decaying weights. It also nests the benchmark predictability regression in Equation 1 with $\alpha_{n,h} = 0$. Beyond its simplicity, this reduced-form recursive representation is also consistent with a non-markovian dynamics for yields where the parameters $\alpha_{n,h}$ and $\gamma_{n,h}$ are functions of the underlying dynamics parameters. The coefficient $1 - \alpha_{n,h}$, with $0 \leq \alpha < 1$, measures the weight on the current forward rates, $\gamma'_{n,h}f_t$, while $\alpha_{n,h}$ controls the weights on past forward rates, via the recursive term, $\mathcal{R}_{t-1,h}$.

Equation 5 can be estimated easily via least-square given the initial value $\mathcal{R}_{q,n,h}^{(n)} = \mathcal{R}_{h}^{(n)}$, which we set to the unconditional population mean. Table 2 presents the results obtained when estimating Equation 5 for each maturity and each holding period separately. Panel (A) displays the $R^2$s. Adding lagged forward rates increase returns predictability across every horizon and bond maturities. The contribution of lagged forward rates is clear. The predictability almost doubles for the 1-month and 2-month horizons, from 3% to 6% and from 6% to around 11%, respectively, and increase substantially at longer horizons. Panel (B) reports estimates of $\alpha_{n,h}$. The estimates exhibit a gradual decline across horizons, ranging between 0.8 and 0.6. Hence, information from past forward rates plays a more important role at shorter maturities.

The results also suggest that the returns forecasting factors exhibit a factor structure across maturities and across horizons. Estimates of $\alpha_{n,h}$ are remarkably similar across maturities (keeping the horizon fixed) and estimates of $\gamma_{n,h}$ exhibit the familiar tent-shaped pattern (unreported). We check the single-factor restriction formally via the following model:

\[ x_{t,h}^{(n)} = b_{n,h} ((1 - \alpha_h)\gamma'_{h}f_t + \alpha_h\mathcal{R}_{t-1,h}) + u_{t,h}^{(n)}, \]

\[ \mathcal{R}_{t,h} = \alpha_h\mathcal{R}_{t-1,h} + (1 - \alpha_h)\gamma'_{h}f_t, \]  

(6)

where the coefficients $\gamma_h$ and $\alpha_h$ do not vary with the maturity. As in CP we impose that $H^{-1}\sum_{h=1}^{H} b_{n,h} = 1$ to identify $b_{n,h}$ from $\gamma_h$ and $\alpha_h$. Panel (A) of Table 3 reports the $R^2$s. Again, the single-factor restriction leads to no predictability loss. The estimates of $\alpha_h$ are also essentially unchanged relative to the unrestricted case. As expected, Panel (C) shows that estimates of the loadings $b_{n,h}$ increases with the maturity of the bond. Panel (D) reports estimates of $\gamma_h$. The tent-shape pattern is clearly apparent and strikingly similar across horizons even if estimation was carried out independently for each horizon. Again, a few linear combinations of forward rates (and its lags) drive bond returns across different holding horizons. Table 4 displays
results from a principal components analysis (PCA) of the returns-forecasting factors, \( R_{t,h} \), across different holding horizons. We find that the first principal component explain 97% of total variations with loadings that spreads out evenly across maturities.

### 1.3 Reconciling with Yields Dynamics

CP shows that a VAR(1) on yields at the *monthly* frequency is inconsistent with the predictability in annual returns. In addition, a VAR(1) model cannot reproduce the added predictability of lagged forward rates by construction: current forward rates must summarize all the predictability within a VAR(1). We introduce the family of Conditional Mean models formally in Section 2, but we first document that simply adding a moving-average component captures the evidence in annual returns.\(^5\)

Following CP, define the yield vector, \( Y_t = [y_t^{(1)} \ y_t^{(2)} \ y_t^{(3)} \ y_t^{(4)} \ y_t^{(5)}]' \) where \( y_t^{(n)} \) is the \( n \)-year zero-coupon yield, and with the following \( \mathbb{P} \)-dynamics:

\[
Y_t = \mathcal{E}_{Y,t-1} + \Sigma_Y \epsilon_{Y,t} \\
\Delta \mathcal{E}_{Y,t} = K_{Y,0} + K_{Y,1} \mathcal{E}_{Y,t-1} + \Sigma_{\mathcal{E}_Y} \epsilon_{Y,t},
\]

where \( \mathcal{E}_{Y,t} \equiv E_t[Y_{t+1}] \) and \( \epsilon_t \) is white noise. For our present purpose, it suffices to note that forecasts from Equation 7 depend on the history of yields and these forecasts are given by an infinite sum similar to that in Equation 5. Hence, Equation 7 has the potential to match the predictability results above. Following CP, we estimate each model and derive the implied population \( R^2 \)'s for the regression in Equation 1. For comparison, and as in CP, we also consider a markovian VAR(1) model and a VAR(12) model.

Panel (A) of Table 5 displays the results. Consistent with CP’s results, the predictability of annual returns implied by the VAR(1) model is typically half that obtained from direct regressions. On the other hand, both our model and the VAR(12) imply population \( R^2 \)'s that are close to the regression results. Therefore, our parsimonious specification can bridge the temporal aggregation gap between the monthly frequency at which the time-series model is estimated and the annual frequencies at which the predictive regression are directly estimated. The VAR(12) is much less parsimonious and the added apparent predictability could result from over-fitting. To check that the predictability gain implied by the VAR(12) is an artefact of sampling

---

\(^5\)Cochrane and Piazzesi (2005) suggest that the importance of lags is due to the presence of small measurement errors in forward rates and that the pattern of coefficients on lagged forward is consistent with the result from a Kalman filter. We consider this channel explicitly in Section 4.
uncertainty, we run an out-of-sample forecast horse race. Starting in the middle of the sample, we estimate the model anew each month and produce forecasts of annual excess returns. Panel (B) of Table 5 displays the forecast RMSEs. Out-of-sample forecasts improve when moving from the VAR(1) to our markovian specification but there is no additional gain when moving to the VAR(12). Hence, the additional predictability from the VAR(12) model does not survive the out-of-sample test and may be due to over-fitting.

2 Conditional Mean DTSM

2.1 Risk-Neutral Dynamics

Dynamic Term Structure Models (DTSMs) of the term structure postulate a small number of $\mathcal{N}$ of risk factors $Z_{t+1}$ driving the cross-section of bond yields. The solution for yields can be derived explicitly whenever the risk factors’ dynamic is tractable across different horizons under the risk-neutral measure $Q$. Hence, a common approach takes a stationary dynamics with Gaussian innovations for $Z_{t+1}$, leading to the class of gaussian DTSM. In addition, a natural approach taken in almost all existing models assumes that $Z_t$ has a Markovian dynamics. Indeed, Proposition 8 clarifies that $Z_t$ must remain Markovian under the risk-neutral measure $Q$ in the class of linear gaussian models.

**Proposition 1** Let $Z_t \in \mathbb{R}^\mathcal{N}$ with stationary conditional Gaussian distribution. If the cross-section of $n$-period yields, $Y^{(n)}_t$, can be expressed as a linear function of $Z_t,$

$$Y^{(n)}_t = A_n + B'_n Z_t,$$  

then $Z_t$ has a markovian dynamics under $Q$ and we can write:

$$\Delta Z_t = K^Q_0 + K^Q_1 Z_{t-1} + \Sigma_Z \epsilon^Q_{Z,t},$$  

where $\epsilon^Q_{Z,t}$ is a standard Gaussian innovation.

This result should not come as a surprise. It is related (but differ) from results in the literature with unspanned macro factors (Duffee, 2011; Joslin, Priebsch, and Singleton, 2010). Nonetheless, Proposition 1 is useful to clarify that once we fix $\mathcal{N}$ the number of linear risk factors required to explain the cross-section of yields, then these
factors must have a linear markovian dynamics under the risk-neutral dynamics. The result is not trivial but depends crucially on the absence of arbitrage opportunity.

The relationship between the loadings $A_n$ and $B_n$ in Equation 8 and the $Q$-dynamics in Equation 9 is standard and given by the no-arbitrage price of zero-coupon bonds,

$$D_{t,n} = E_t^Q [e^{-\sum_{i=0}^{n-1} r_{t+i}}] = e^{A_n + B'_n Z_t}, \quad (10)$$

for $n > 1$, where $A_n$ and $B_n$ satisfy the first-difference equations:

$$A_{n+1} - A_n = K^Q_0 B_n + \frac{1}{2} B'_n \Sigma_Z \Sigma'_Z B_n - \rho_0$$
$$B_{n+1} - B_n = K^Q_1 B_n - \rho_1, \quad (11)$$

where the coefficients for yields are $A_n = -A_n/n$ and $B_n = -B_n/n$, and where $\rho_0$ and $\rho_1$ are the loadings on the short rate:

$$r_t = \rho_0 + \rho_1 Z_t. \quad (12)$$

### 2.2 Historical Dynamics

Proposition 2 introduces the family of Conditional Mean Gaussian Dynamic Term Structure Model (or Conditional Mean GDTSM) and exhibits the pricing kernel that is consistent with the $Q$-dynamics in Equation 11.

**Proposition 2** Let $Z_t \in \mathbb{R}^N$ with the following generic $\mathbb{P}$-dynamics:

$$Z_t = \mathcal{E}^\mathbb{P}_{Z,t-1} + \Sigma_Z \epsilon^\mathbb{P}_t$$
$$\Delta \mathcal{E}^\mathbb{P}_{Z,t} = K^\mathbb{P}_0 + K^\mathbb{P}_1 \mathcal{E}^\mathbb{P}_{Z,t-1} + \Sigma_{\epsilon_Z} \epsilon^\mathbb{P}_t, \quad (13)$$

where $\mathcal{E}^\mathbb{P}_{Z,t} \equiv E^\mathbb{P}_t [Z_{t+1}]$ and $\epsilon^\mathbb{P}_t$ is a standard Gaussian white noise. Let $M_{t+1}$, the pricing kernel given by:

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

\[ \lambda_t \equiv (\Sigma_Z \Sigma_Z')^{-1} \left( \lambda_0 + \lambda_1 Z_t + \lambda_2 \mathcal{E}_{Z,t-1}^P \right), \]  \hspace{1cm} (14)

then the $\mathbb{Q}$-dynamics of $Z_t$ is given by Equation 11. The unique mapping between parameters under $\mathbb{P}$ and $\mathbb{Q}$ is given by:

\[
\begin{align*}
K_P^0 &= \lambda_0 + K_Q^0 \\
K_P^1 &= \lambda_1 + \lambda_2 + K_Q^1 \\
\Sigma_{\varepsilon_Z} &= (\lambda_1 + K_Q^1 + I_N) \Sigma_Z.
\end{align*}
\]  \hspace{1cm} (15)

Several comments are in order. First, the conditional expectations $\mathcal{E}_{Z,t}^P$ follow a VAR(1) dynamics under $\mathbb{P}$ while the standard model specifies a VAR(1) for $Z_t$. In the latter case, we have that $\mathcal{E}_{Z,t}^P$ is a function of $Z_t$. Hence, our approach breaks the mapping between $Z_t$ and $\mathcal{E}_{Z,t-1}$ inherent in any markovian model. Instead, $\mathcal{E}_{Z,t}^P$ is a function of $Z_t$ and of the entire history of $Z_t$. This arise in equilibrium models where $Z_t$ is part of a broader equilibrium involving other financial or macro variables, say $\Xi_t \equiv [Z_t, Z_{2,t}]$, where $\Xi_t$ 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.\(^6\)

Second, the pricing kernel is standard with but prices of risk that are functions of both $Z_t$ and $\mathcal{E}_{Z,t-1}$-1. The standard case is nested with $\lambda_2 = 0$ whereas the prices of risk only depend on the current value $Z_t$. Hence, extending the dependence of the prices of risk on both $Z_t$ and $\mathcal{E}_{Z,t-1}$ is equivalent to a conditional mean specification of yields under the $\mathbb{P}$ measure.\(^7\)

Third, note that the conditional dynamics of $Z_t$ and $\mathcal{E}_{Z,t}^P$ are driven by the same $\mathcal{N}$ shocks $\epsilon_t^P$. Hence, while the number of state variables has increased relative to the standard case (from $\mathcal{N}$ to $2\mathcal{N}$) there has a been no increase in sources of risk.\(^8\)

\(^6\)An alternative is to include lags of $Z_t$ in a VAR(p) representation. We argue that the conditional mean representation is more intuitive in the context of term structure models (where conditional expectations play a key role) and, in addition, that the alternative VAR(12) approach is not parsimonious. For instance, Ang, Piazzesi, and Wei (2006) uses a VAR(12) dynamics.

\(^7\)A pricing kernel that involves long lags of the state variables is consistent with habit specification or with MA components in the dynamics of the state variables in long-run risk economies (e.g., see the NBER working paper version of Bansal and Yaron (2004)).

\(^8\)In the absence of measurement errors, $\mathcal{N}$ corresponds to the number of principal components in the cross-section of yields. Note that $\mathcal{N}$ corresponds to number of yield factors that are required to explain the cross-section of yields by construction.
The innovations to $Z_t$ are given by $\Sigma_Z \epsilon_t^P$, while the innovations to $\varepsilon_{Z,t}^P$ are given by $\Sigma_Z \epsilon_t^P$. Therefore, the covariance matrix for $Z_t$ and $\varepsilon_{Z,t}^P$, taken jointly, has rank $N$ only (conditionally or unconditionally as estimated via principal component analysis, say).

Fourth, and finally, Equation 13 corresponds to the conditional mean representation of an unrestricted VARMA(1,1).\(^9\) This representation was introduced in Fioren-tini and Sentana (1998) in the context of time-series models: the conditional expectations depend on lagged expectations, via an auto-regressive term, and on the current shock to the observable, analogous to a GARCH model for the conditional variance. In particular, this implies that $\varepsilon_{Z,t}^P$ is not latent but can be filtered from the time-series of $\varepsilon_{Z,t}^P$ given some initial values $\varepsilon_{Z,0}^P$. For future reference, the standard VAR(1) model is nested with

$$
\Sigma_Z = (I_N + K_1^P) \Sigma_Z,
$$

in which case we have that $\varepsilon_{Z,t}^P = K_0^P + (I_N + K_1^P) Z_t$ and that $Z_t$ is markovian.

### 2.3 Canonical Form

We develop a canonical form for the class of conditional mean DGTSM, which can be seen as a direct extension of the standard JSZ canonical form. First, we extend Proposition 1 of JSZ to the case where the risk factors have non-markovian $\mathbb{P}$-dynamics.

**Proposition 3** Every canonical conditional mean GDTSM is observationally equivalent to the canonical model with $\nu_t = \nu \cdot Z_t$,

$$
\begin{align*}
Z_t &= \varepsilon_{Z,t-1}^P + \Sigma_Z \epsilon_t^P, \\
\Delta \varepsilon_{Z,t}^P &= K_0^P + K_1^P \varepsilon_{Z,t-1}^P + \Sigma_Z \epsilon_t^P, \\
\Delta Z_t &= K_0^Q + K_1^Q Z_{t-1} + \Sigma_Z \epsilon_{Z,t}^Q
\end{align*}
$$


where $\nu$ is a vector of ones, $\Sigma_Z$ is lower triangular with positive diagonal elements, $K_1^Q$ is an ordered real Jordan form, $K_0^Q = k_{\infty}^Q$, $K_{0,i}^Q = 0$ for $i \neq 1$, and both $\epsilon_{Z,t}^Q$ and...\(^{10}\)
As in JSZ the $Q$-dynamics of the risk factors $Z_t$ is parameterized by an autoregressive matrix $K^Q_1$ in Jordan form (see JSZ) and $k^Q_{\infty}$. On the other hand, the $P$-dynamics of the risk factors is free. Hence, this result exactly matches Proposition 1 in JSZ whenever the restriction $\Sigma_{\epsilon Z} = (I_N + K^Q_1)\Sigma_{\epsilon Z}$ holds or, equivalently, whenever $\lambda_2 = 0$.

Next, we let $(m_1, m_2, \ldots, m_J)$ be the set of maturities (in years) of the bonds used in estimation with $J > N$ and $Y_t \equiv (y_{t,m_1}, \ldots, y_{t,m_J}) \in \mathbb{R}^J$ be the set of model-implied yields (see Equation 8) and $Y_t^o$ the corresponding set of observed yield (possibly observed with errors). Suppose that $N$ portfolios $P_t$ of yield with weights $W \in \mathbb{R}^{N \times J}$ are observed without error. Theorem 1 formally establishes our canonical form for the class of conditional mean DGTSM.

**Theorem 1** If the portfolios $P_t \equiv WY_t^o$ are observed without error, then any canonical conditional mean GDTSM is observationally equivalent to a unique conditional mean GDTSM whose risk factors are $P_t$. Moreover, the $Q$-distribution of $P_t$ is uniquely determined by $(\lambda^Q, k^Q_{\infty}, \Sigma_P)$, where $\lambda^Q$ is ordered. That is

\[
\begin{align*}
\mathcal{P}_t &= \mathcal{E}_{P,t-1}^P + \Sigma_P \epsilon_{P,t}^P \\
\Delta \mathcal{E}_{P,t}^P &= K^P_0 + K^P_{1} \epsilon_{P,t-1}^P + \Sigma_P \epsilon_{P,t}^P \\
\Delta \mathcal{P}_t &= K^Q_0 + K^Q_1 \mathcal{P}_{t-1} + \Sigma_P \epsilon_{P,t}^Q \\
\mathcal{R}_t &= \rho_{0,p} + \rho_{1,p} \cdot \mathcal{P}_t
\end{align*}
\]

is a canonical conditional mean GDTSM where $\mathcal{E}_{P,t}^P \equiv E_t^P[P_{t+1}]$, and where $K^Q_0, K^Q_1, \rho_{0,p}, \rho_{1,p}$ are explicit functions of $(\lambda^Q, k^Q_{\infty}, \Sigma_P)$. The canonical form is parameterized by $\Theta^P = (\lambda^Q, k^Q_{\infty}, \Sigma_P, K^P_0, K^P_1, \Sigma_{\epsilon P})$.

We refer to the conditional mean GDTSM in Theorem 1 as the FF canonical form parameterized by $\Theta^P$. The parameters $(K^Q_0, K^Q_1, \rho_{0,p}, \rho_{1,p})$ are given by:

\[
\begin{align*}
K^Q_0 &= DK^Q_0 - K^Q_{1,p}C \\
K^Q_1 &= DJ (\lambda^Q) D^{-1} \\
\rho_{0,p} &= -t'D^{-1}C \\
\rho_{1,p}^t &= t'D^{-1}
\end{align*}
\]

with $K^Q_{0,i} = k^Q_{\infty}$, $K^Q_{0,i} = 0$ for $i \neq 1$, $C = A_P (\Theta^Q)$ and $D = B_P (\Theta^Q)'$ given the portfolios loadings,

\[
\mathcal{P}_t = A_P (\Theta^Q) + B_P (\Theta^Q)' Z_t.
\]
The FF canonical implies that the parameters governing the dynamics of $P_t$ can be estimated separately from the parameters governing the likelihood of the observed yields (as in JSZ). Suppose that the measurement errors $Y_t^o - Y_t$ have conditional distribution $P^{\theta_m}$, for some parameter $\theta_m \in \Theta_m$, and that these errors are independent of lagged measurement errors.\(^{11}\) Then the conditional likelihood of the observed data $Y_t^o$ is

$$f(Y_t^o|\mathcal{E}_{P_t,t-1}; \Theta) = f(Y_t|\mathcal{P}_t; \lambda^Q, k^Q, \Sigma_p, P^{\theta_m}) \times (\mathcal{P}_t|\mathcal{E}_{P_{t-1},t-1}; K^P_{0,p}, K^P_{1,p}, \Sigma_{\mathcal{E}_p}, \Sigma_p), \quad (20)$$

where one can obtain consistent estimates of $K^P_{0,p}$, $K^P_{1,p}$, and $\Sigma_{\mathcal{E}_p}$ based on the second component of the likelihood using observations of $\mathcal{P}_t$. For instance, the historical dynamics of $\mathcal{P}_t$ can be estimated freely based via MLE. Then, one can used the likelihood for $Y_t^o$ and the estimate for $\Sigma_p$ to estimate the remaining parameters.

### 2.4 History-dependent bond returns

Section 1 provides evidence that bond risk premium cannot be summarized by current yield observation, and that bond premiums can be well-approximated by a recursive structure (like that in Equation 6). Under the conditions of Proposition 1, the excess returns from holding an $n$-period bond between $t$ and $t + h$ is given by:

$$x^{(n)}_{t,h} = -\frac{h}{2} B'_{n-1,p} \Sigma_p \Sigma' \Sigma_p B'_{n-1,p} + B'_{n-1,p} \sum_{j=1}^{h} (\Delta \mathcal{P}_{t+j} - E^Q_{t+j-1} [\Delta \mathcal{P}_{t+j}]), \quad (21)$$

where $B_{n-1,p}$ is defined analogously to $B_{n-1}$ in Equation 11 but where the portfolios $\mathcal{P}_t$ are the risk factors. Taking conditional expectations under the $\mathbb{P}$ measure, each summand in the second term becomes:

$$E^P_t \left[-K^Q_{0,p} + (I_N - K^Q_{1,p}) \Delta \mathcal{P}_{t+j} \right], \quad (22)$$

since the time-$t$ conditional expectation operator under $\mathbb{Q}$ is a function of $\mathcal{P}_t$ only. Therefore, Proposition 1 implies that the predictability of bond excess returns in Equation 21 depends on the history of forward rates if and only if the $\mathbb{P}$-expectation of $\mathcal{P}_t$ is non-markovian. In other words, markovian models imply that the bond risk premium does not involves the history of $\mathcal{P}_t$. Proposition 4 makes this point precise in the context of conditional mean GDTSM.

**Proposition 4** For conditional mean GDTSM, the risk premium from holding an

---

\(^{11}\)As in JSZ, the measurement errors must also satisfy the consistency condition $\mathbb{P}(WY_t^o = \mathcal{P}_t) = 1.$
The $n$-period bond between $t$ and $t + h$,

\[ brp_{t,h}^{(n)} \equiv E_t[x_{t+h}^{(n)}], \]  

is given by:

\[ brp_{t,h}^{(n)} = \delta_{h,0} + \delta_{h,1}^p \mathcal{P}_t + \delta_{h,2}^p \mathcal{E}_{\mathcal{P},t-1}, \]  

where $\delta_{h,2} = 0$ if and only if $\mathcal{P}_t$ is markovian under $\mathbb{P}$.

Proposition 4 confirms the observation made above that the role of yields’ history in the $\mathbb{P}$-dynamics and the role of history in the bond risk premium are intertwined. In addition, the bond risk premium in Equation 24 is similar to the reduced-form specification in Equation 5. The risk premium depends on the current forward rates, via the yield portfolios $\mathcal{P}_t$, but also depends on past forward rates, via $\mathcal{E}_{\mathcal{P},t-1}$ with a decaying pattern for the coefficients. Proposition 4 justifies our specification to capture the bond returns predicability (See Table 2 and Table 5).

3 Results

3.1 Data and Estimation

Estimation is based on end-of-month CRSP data from December 1963 until December 2007. We use zero coupon yields with maturities of 3 months, 6 months, and then from 1 to 5 years. (Longer maturities are not available in the early part of the sample.) This section compares models with $\mathcal{N} = 5$ factors. This choice of sample period, of maturities and of $\mathcal{N}$ eases the comparison with results from Duffee (2011). (See Section 4 below.) Similarly, CP emphasize that a markovian model with $\mathcal{N} = 5$ factors sampled at the monthly frequency cannot recover the predictability evidence from direct regression at the annual horizons. In line with CP, we set $\mathcal{P}_t = [f_t^{(1)} f_t^{(2)} f_t^{(3)} f_t^{(4)} f_t^{(5)}]'$, where $f_t^{(n)}$ the one-year forward rate $n - 1$ year ahead. We estimate a VAR(1) model and a non-Markovian Conditional Mean model, labelled the $VAR_5$ and $CM_5$, respectively. We estimate parameters of the $\mathbb{P}$-dynamics via Maximum Likelihood and we estimate parameters of the $\mathcal{Q}$-dynamics separately by minimizing the squared pricing errors of the remaining yields. Future work will consider cases where $\mathcal{N} < 5$. 

15
3.2 Predictability

3.2.1 Variance Ratios

We compare the variability of bond risk premiums from each model. The current cross-section of yields determines the risk premium within the \( VAR_5 \) model. In contrast, Equation 24 shows that the bond risk premium also depends on the history of yields in the \( CM_5 \) model. How large is the role of history? One measure is the ratio of the bond risk premium variance from the \( VAR_5 \) model relative to the \( CM_5 \).

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. Then, we compute the sample variance for each maturity and for each horizon.

Table 7A reports the ratio of the sample variances from each model. A value less than one indicates that the bond risk premium from the \( VAR_5 \) is less variable than the risk premium from the \( CM_5 \) model, and provides a gauge of the contribution of past yields to the forecast. This contribution is large. The ratios typically hover around 40-50\%, except in cases where the returns horizon close to the bond maturity. For instance, the ratio increases from 32\% up to 55\% and from 18\% up to 59\% for the 2-year and the 5-year bond, respectively. The ratio for the 6-month and the 1-year bond is also low at short horizons, 52\% and 41\% respectively, but rises rapidly when we reach horizons near the bond maturity. Overall, the history of yields generate significant risk premium variability that is not contained in current yields.

3.2.2 Returns Forecast Errors

In a second step, we ask whether the greater bond risk premium variability produces greater accuracy in forecasting excess returns. We compute the sample variance of the excess returns forecast errors from the following Mincer-Zarnowitz regressions,

\[
x_{r_{t,h}}^{(n)} = a + b \times brp_{t,h}^{(n)} + u_{t,h}^{(n)}
\]

in each model. Table 7B reports the ratios of the \( R^2 \)s in the \( VAR_5 \) model relative to the \( CM_5 \) model. A value less than one indicates that the bond risk premium from the \( VAR_5 \) are less accurate than the risk premium from the \( R^2 \)s in the \( CM_5 \) model, and provides a gauge of the contribution of past yields to the forecast accuracy. Again, this contribution is large. In fact, the ratios in Panel (B) are close to the ratios in Panel (A): much of the added variability in the risk premium translates into an improved accuracy. The relative improvements are striking at the shorter horizons.
(in part because the predictability implied by the $VAR_5$ is so low) but remains large across the board. For instance, in the case of the 2-year and the 5-year bonds, using only the information contained in current yields produces $R^2$s that are between 13% and 73% and between 27% and 64% of the $R^2$s based on the $CM_5$. Relative improvements that are higher for the shorter horizons are also consistent with the evidence from direct regressions in Table 1, 2 and 3.

The point estimates for $b$ in Equation 25 vary substantially but they are typically not significantly different from 1. In addition, the coefficient estimates from the $VAR_5$ and from the $CM_5$ are typically close to each other, and the gain from using the information contained in past yields are not attributable to different point estimates in Equation 25. One simple way to check this is to compare the $R^2$s from the unrestricted predictive regressions with the $R^2$s in the restricted case:

$$x_{r_t,h}^{(n)} = b r_{t,h}^{(n)} + u_{t,h}^{(n)}.$$  

The null that $a = 0$ and $b = 0$ is consistent with the underlying term structure models. Table 7C reports the ratios of the $R^2$s in the restricted regressions relative from unrestricted regression, using the bond risk premium from the $CM_5$ model in each case. With a few exceptions, the ratios are uniformly close to 1. This confirms that the predictability gains follows from conditioning on the history of yields.

Finally, how much of the predictability captured by the recursive regression in Section 1 is also captured in the model-implied the risk premium? Consider a regression of the predicted excess returns from Equation 3 on the model risk premium,

$$\hat{x}_{t,h}^{(n)} = a + b \times r_{t,h}^{(n)} + u_{t,h}^{(n)}.$$  

Note that the variable on the left-hand side is measured with errors and the estimates of $b$ may be biased. In addition, $\hat{x}_{t,h}^{(n)}$ has been obtained from a finite sample and may suffer from over-fitting. Hence, it is not clear how high a successful $R^2$s should be in Equation 27. However, what is of importance is how much we gain from relaxing the Markov assumption. One simple check is the ratio of the $R^2$s from Equation 27 obtained with the risk premium from the $VAR_5$ model relative to that obtained from the $CM_5$ model. Figure 2 shows the ratios of the $R^2$s for bonds with 2, 3, 4 and 5 years to maturity and across different horizons up to 12 months. Again, the improvements

---

12This was expected since the coefficient estimates from $b$ are not statistically different than one in almost every case. Note also that we did not report results based on the $R^2$s from the restricted regression based on the bond risk premium from the $VAR_5$ since these are negative in several cases.
are substantial, with ratios ranging between 55% and 85% for 3-month returns and between 65% and 95% for 12-month returns.

3.2.3 CP Regression

The predictability gain documented so far results from breaking the Markov assumption for yields. This single assumption introduces two mechanisms limiting the VAR model’s ability to generate predictability. The most apparent channel restricts the information set to current yields, excluding the history of yields from conditional forecasts. The previous section shows the importance of this channel. A slightly less apparent channel restricts the ability of the VAR model to link the monthly yield dynamics to bond returns at longer horizons. Heuristically, the condition in Equation 16 implies that current shocks must have the same weights than past expectations in updating the forecasts $E_{p,t}$, introducing a tension between short-term forecasts and long-term forecasts. This section shows the importance of this channel for the model’s ability to aggregate forecast over longer horizons.

Comparing the population coefficients in Cochrane-Piazzesi regressions of bond excess returns on forward rates only (as in Section 1) provides one way to see that the effect of time-aggregation is separate from the information content from the history of yield. This criteria is analogous to using the coefficients in the Campbell-Shiller regression coefficients as the benchmark to gauge DTSMs. CP show that the $VAR_5$ does not match these coefficients for annual returns and we find the same result in our longer sample. But does the population coefficients implied by the $CM_5$ model match the regression estimates? In other words, does the $CM_5$ help bridging the aggregation gap between the monthly sampling frequency and returns at longer horizons?

Figure 1 displays the coefficients for bond with maturity of 3 and 5 years and for horizons of 3, 6 and 12 months across Panel (A)-(C). Each panel compares the estimates from a direct regressions of excess returns on forward with the population coefficients implied by the $VAR_5$ and $CM_5$ models. Going through the figures, it is clear that the OLS coefficients display the common tent-shape across horizons. The coefficients from the $CM_5$ model also display the expected tent-shape but coefficients from the $VAR_5$ do not. In addition, coefficients from the $CM_5$ model are much closer to than their sample counterpart (obtained from direct regressions) than population from the $VAR_5$. 

18
3.3 Yield Decomposition

Finally, we assess whether the differences between the bond risk premium predicted by each model translate into meaningful differences between their respective decomposition of yields. In particular, we focus on the expectation component of the 10-year yield – which plays a key role in ongoing discussions surrounding the Federal Reserve policy actions since 2009. For that purpose, we re-estimate the model using data until December 2011. Figure 3 displays the expectation component of the 10-year yield, zooming on the period since 2003.

Both models agree on the behavior of expectations between 2003 and the middle of 2007 and show a gradual increase in advance of the 2004-2006 episodes. Hence, as several other term structure models, Greenspan’s conundrum is explained via declines in risk premium for both models. The $CM_5$ model suggests that expectations are more variables (not as smooth as those implied from the $VAR_5$) but the differences are small overall. Most interestingly, the expectation components from each model exhibit a large difference starting in 2007. This gap widens in the second half of 2007, reaching 0.5%, and persists throughout the rest of the sample. The risk premium implied by the non-markovian model is lower, and this is consistent with the substantial body evidence documenting the effect of quantitative easing by the Fed on the risk premium, as well as the safe haven provided to investors by Treasury securities in volatile markets.

4 Markovian Yields and Measurement Errors

CP suggest one interpretation of the role played by lagged forward rates: “time- $t$ yields (or prices, or forwards) truly are sufficient state variables, but there are small measurement errors that are poorly correlated over time”. Then, it follows from the standard Kalman “that the best guess of the true state is a geometrically weighted moving average.” (See Cochrane and Piazzesi (2005), p.154.) This is also consistent with recent findings in Duffee (2011) who shows that significant share of the information contained in filtered state vector is not spanned by current yields but it is obtained from the history of yields via the Kalman filter. This section studies this rationalization extensively. We find that measurement errors on their own are unlikely to be a sufficient explanation, and that the underlying state are likely to

---

13Both model capture the rapid increase in the expectation toward the end of 2004 and the beginning of 2005 and attribute the conundrum posed by the decline of yields in that period to a lower risk premium. This is consistent with previous evidence.
be non-Markovian themselves. Nonetheless, these two channels cannot be identified separately without more structure or more data.

### 4.1 Conditional mean representation of the Kalman filter

We follow CP’s intuition that measurement errors can generate the desired dynamics. Appendix A.1 considers a model where \( \mathcal{N} \) risk factors follow a Markovian VAR1 dynamics, but where all yields are measured with errors. It follows that any (unobserved) yield portfolios \( \tilde{P}_t \) have a Markovian dynamics:

\[
\begin{align*}
\Delta \tilde{P}_{t+1} &= \mu_w + K_w \tilde{P}_t + \Omega_{w,t+1}^{1/2} \epsilon_{w,t+1} \\
\Delta \tilde{P}_t &= \mu_w + K_w \tilde{P}_t + \Omega_w^{1/2} \epsilon_{w,t+1},
\end{align*}
\]

(28)

where \( \mu_w, K_w \) and \( \Omega_w \) depends on the underlying state dynamics as well as the weight matrix used to form the portfolios. However, the observed yield portfolios follow a conditional mean dynamics (e.g, as in Equation 13):

\[
\begin{align*}
\mathcal{P}_{t+1} &= \mathcal{E}_{\mathcal{P},t} + \eta_{t+1} \\
\Delta \mathcal{E}_{\mathcal{P},t+1} &= \mu_w + K_w \mathcal{E}_{\mathcal{P},t} + (K_w + I_N)P(P + R)^{-1}\eta_{t+1},
\end{align*}
\]

(29)

where \( \mathcal{E}_{\mathcal{P},t} \) is the (optimal) Kalman filtering predictions for \( \tilde{P}_t \):

\[
\mathcal{E}_{\mathcal{P},t} \equiv E[\tilde{P}_{t+1} | \mathcal{P}_1, \ldots, \mathcal{P}_t].
\]

The prediction error \( \eta_{t+1} \) has mean zero and covariance matrix \( P + R \), where \( R \) is the covariance matrix of measurement errors in yields and \( P \) is the steady-state Kalman filter covariance matrix. The matrix \( P \) solves the following algebraic Ricatti equation:

\[
P = (K_w + I_N)[P - P(P + R)^{-1}P](K_w + I_N)' + \Omega_w.
\]

### 4.2 Restrictions implied by the Kalman filter

Importantly, the dynamics of \( \mathcal{P}_{t+1} \) in Equation 29 is a restricted case of the general conditional mean dynamics in Equation 13. A simple count of parameters shows that \((N - 1)N/2\) parameters are not free. But the following is more instructive. In the notation of Theorem 1, we have \( \Sigma_{\mathcal{E}_p} = (K_w + I_N)P(P + R)^{-1} \) and, from Equation 16,
the dynamics is non-Markovian only if \((K_w + I_N)^{-1}\Sigma \xi_p \neq I\). Indeed, we have that

\[
(K_w + I_N)^{-1}\Sigma \xi_p = P(P + R)^{-1}
\]

(30)

where \(P(P + R)^{-1} \neq I_N\) whenever \(R \neq 0\). In other words, the presence of measurement errors generate non-Markovian dynamics via the Kalman filter.

However, \(P\) and \(R\) are covariance matrices, which implies that the left-hand side is a positive definite matrix. Therefore, \((K_w + I_N)^{-1}\) and \(\Sigma \xi_p\) are not entirely free from each other. Departures from the Markovian case become more restricted as more structure is imposed on the covariance matrix \(R\). For instance, two common assumptions are that \(R\) is a diagonal matrix or, as in Duffee (2011) that \(R = \sigma^2 I\). Finally, note that we obtain the unrestricted conditional mean model if we loosen the interpretation in terms of measurement errors and allow \(R\) to be an arbitrary matrix (which eliminates all the \((N-1)N/2\) restrictions).

### 4.3 The role of measurement errors

Hence, the analysis shows that the non-markovian dynamics generated from the Kalman filter is a restricted case of the non-markovian dynamics in Theorem 1. The matrix \(R\) contains all the restrictions that followed from an interpretation in terms of measurement errors. The magnitude of the matrix \(R\) also helps answer the question: How far is the restricted model from the unrestricted model? Answering this question exactly requires that we solve the Ricatti equation for \(P\) in closed-form, which is not in general feasible. However, we derive a first-order approximation around \(R = 0\), where \(P(P + R)^{-1} = I_N\) and we have a Markovian case. Appendix A.2 shows that, to a good approximation,

\[
(I_N - P(P + R)^{-1}) \approx R\Omega_w^{-1},
\]

(31)

for \(R\) close to 0, or equivalently,

\[
(K_w + I_N)^{-1}\Sigma \xi_p \approx R\Omega_w^{-1}.
\]

(32)

Therefore, the dynamics implied from the Kalman filter will be “close” to Markovian if \(R\) is “small” relative to \(\Omega_w^{-1}\). That is, if the variability of the measurement errors is small relative to the variability of the innovations to unobserved yield portfolios \(\mathcal{P}_t\).
4.4 Empirical Results

We estimate different version of Equation 29 via MLE. We consider the following cases:

1. $CM_5$: $R$ is free,
2. $CM_5 - KF1$: $R$ is positive definite,
3. $CM_5 - KF2$: $R$ a diagonal matrix, and
4. $CM_5 - KF3$: $R = \sigma^2 I$,

and the $VAR_5$ model corresponds to the case with $R = 0$.

Variance ratios provide one simple way to compare the effect of each successive restriction. Similar to Table 6, Table 7 shows the ratios of the sample bond risk premium variance from the $VAR_5$ relative to each variant of the KF model. The risk premium variability of the $VAR_5$ is between 90% and 100% of the variability obtained from the $CM_5 - KF3$, where $R = \sigma^2 I$. The case where $R$ has a free diagonal produces more variability for some maturities. The variability from the $VAR_5$ now range between 73% and 96% of the variability obtained in the $CM_5 - KF2$. Relaxing the $R$ matrix further to a positive-definite and full covariance matrix produces substantial gain, and the the added variability from the $CM_5 - KF1$ is closed to the variability implicit in the unrestricted $CM_5$.

However, an interpretation of the estimated $R$ matrix in terms of measurement errors appears unreasonable. In the case where $R = \sigma^2 I$, the measurement errors would have a standard deviations of 15 basis points (annualized). This seems too high relative to alternative estimates of the magnitude of measurement errors. For example, Duffee (2011) estimates the standard deviation of measurement errors of 5 basis points.\footnote{Estimates in Duffee (2011) are based on the joint likelihood of the time-series of yields and of the cross-section of yields.} Similarly, in the case where $R$ is a diagonal matrix, the highest standard deviation reaches 25 basis points. Finally, the case where the matrix $R$ is an arbitrary covariance matrix imply correlations between would-be yield measurement errors ranging between -0.8 and 0.8.

Table 7 also reports the number of added parameters and the associated likelihood gains from the $VAR_5$ to the $CM_5 - KF3$, and then from the $CM_5 - KF3$ to the $CM_5 - KF2$, from the $CM_5 - KF1$ to the $CM_5 - KF1$, and, finally from the
$CM_5 - KF1$ to the $CM_5$. In each case, the restricted model is rejected relative to the nearest unrestricted model for any common level of statistical significance. The first implication is that, as in Duffee (2011), measurement errors (inevitably) play a significant role and introduce a role for the time-series dynamics to reveal factors that are hard to perceive from the contemporaneous cross-section of yields. However, the second implication is that measurement errors on their own do not generate all the evidence of non-Markovian dynamics apparent in the data.

5 Conclusion

This draft develops the class of Conditional Mean GDTSM based on a conditional mean representation of the yield factors. Yields have non-markovian dynamics under the historical measure $P$, but yields have an affine Markovian dynamics under the risk-neutral measure $Q$, consistent with no-arbitrage. Empirically, this approach captures much of the predictive content of forward rates for bond excess returns. It also generate the patterns of coefficient in Cochrane-Piazzesi regression. Both facts stand as challenges to standard Markovian models. Our results extends those in Dai and Singleton (2002), showing that an affine but non-markovian model can reproduce the reduced-form evidence of bond returns predictability. We show that these differences at times have substantial effects on the decomposition of yields. In particular, the decomposition from a non-markovian model implies a lower risk premium and a higher expectation component, by 0.5%, from 2008 until 2011, when our sample ends.

Importantly, as pointed out in Cochrane and Piazzesi (2005) and documented in Duffee (2011), the combination of Markovian dynamics and of measurement errors in yields generate non-Markovian dynamics for the filtered states – and thus for predicted yields – via the Kalman filter. We show that the Conditional Mean GDTSM nests the population dynamics of the filtered dynamics. We also show that while measurement errors are likely to participate in the observed non-Markovian dynamics, they cannot on their own generate the pattern of predictability.

However, much remain to be done. Beyond a range of robustness and econometric checks, including out-of-sample exercise, we have yet to consider models with a lower number of risk factors. We must also re-visit the potential connection between these new estimates of the risk premium and macroeconomic data. Finally, future research could bring data on bid-ask spread to separately identified the effect of measurement errors from the effect of non-Markovian risk factors.
References


A Appendix

A.1 Conditional Mean Representation of the Kalman Filter

To fix ideas, we consider the specification in Duffee (2011), but with slight changes in the notation: the unobserved state vector $X_t \in \mathbb{R}^N$ has Markovian dynamics given by

\begin{align*}
X_{t+1} &= \mu_x + \phi_x X_t + \Omega^{1/2}_x \epsilon_{x,t+1} \\
X_{t+1} &= \phi^Q_x X_t + \Omega^{1/2}_x \epsilon^Q_{x,t+1},
\end{align*}

(33)

where $\mu_x = \lambda_{x0}$ and $\phi_x = \phi^Q_x + \lambda_{x0}$, and the loadings on the short rate are given by

\begin{equation}
\begin{aligned}
r_t &= \delta_{x0} + \delta_{x1} X_t.
\end{aligned}
\end{equation}

(34)

See, for instance, the parameter estimates in Table 1 of Duffee (2011). Note the subscript $x$ to distinguish from our notation. The true unobserved yields are given by

\begin{equation}
\begin{aligned}
\tilde{y}^{(n)}_t &= a_n + b'_n X_t
\end{aligned}
\end{equation}

(35)

with coefficients given by standard recursions. The observed yields are measured with errors:

\begin{equation}
\begin{aligned}
y^{(n)}_t &= \tilde{y}^{(n)}_t + \eta^{(n)}_t,
\end{aligned}
\end{equation}

(35)

where $\eta^{(n)}_t$ has covariance matrix $R$ ($R = \sigma^2 I$ in Duffee (2011)). Consider the unobserved portfolios of yields defined via the $N \times N$ matrix $W$:

\begin{equation}
\begin{aligned}
\tilde{P}_t &= W \tilde{Y}_t = WA + WBX_t = A_w + B_w X_t,
\end{aligned}
\end{equation}

for some appropriately defined coefficients $A_w$ and $B_w$, and where $\tilde{Y}_t$ stacks $N$ yields. These unobserved portfolio also has Markovian dynamics:

\begin{align*}
\tilde{P}_{t+1} &= \mu_w + \phi_w \tilde{P}_t + \Omega^{1/2}_w \epsilon_{w,t+1} \\
\tilde{P}_{t+1} &= \mu^Q_w + \phi^Q_w \tilde{P}_t + \Omega^{1/2}_w \epsilon^Q_{w,t+1},
\end{align*}

(36)

where $\epsilon_{w,t+1}$ and $\epsilon^Q_{w,t+1}$ are i.i.d. standard Gaussian innovations, with the following parameters:

\begin{align*}
\phi_w &= B_w \phi_x B_w^{-1} & \mu_w &= A_w + B_w \mu_x - \phi_w A_w \\
\phi^Q_w &= B_w \phi_x B_w^{-1} & \mu^Q_w &= A_w - \phi^Q_w A_w,
\end{align*}

(37)

(38)

and where $\Omega_w = B_w \Omega B'_w$. The observed portfolios are contaminated with measurement errors:

\begin{equation}
\begin{aligned}
P_t &= \tilde{P}_t + \eta_{w,t}.
\end{aligned}
\end{equation}

(39)

The $N \times 1$ vector of measurement errors $\eta_{w,t}$ has variance $WRW'$. This model can be estimated via the Kalman filter for an arbitrary non-singular matrix $W$ (the case in Duffee (2011) has $W = I$), yielding the filtered estimates of $\tilde{P}_t$:

\begin{equation}
\begin{aligned}
\mathcal{E}_{\tilde{P},t} &= E[\tilde{P}_{t+1} | P_t, \ldots, P_1],
\end{aligned}
\end{equation}

26
which is the analog to $\hat{P}_{t+1\mid t}$ in Kalman filter notation. Then, the observed portfolios have the following Conditional Mean dynamics in terms of the filtered estimates,

$$
\begin{align*}
\mathcal{P}_{t+1} &= \mathcal{E}_{\mathcal{P},t} + \eta_{t+1} \\
\mathcal{E}_{\mathcal{P},t+1} &= \mu_w + \phi_w \mathcal{E}_{\mathcal{P},t} + \phi_w G_t \eta_{t+1}
\end{align*}
$$

(40)

where $G_t$ is the Kalman gain and $\eta_{t+1} \sim N(0, \Omega_{t,P})$. In terms of the standard Kalman notation (see e.g., Hamilton (1994)), the matrix $G_t$ and $\Omega_{t,P}$ are given by:

$$
\begin{align*}
\Omega_{t,P} &= H'P_{t\mid t-1}H + R \\
G_t &= P_{t\mid t-1}H (\Omega_{t,P})^{-1},
\end{align*}
$$

where the matrix $P_{t\mid t-1}$ is given via the following recursion,

$$
\begin{align*}
P_{t\mid t-1} &= FP_{t-1\mid t-1}F' + Q \\
P_{t\mid t} &= (I - G_t H') P_{t\mid t-1},
\end{align*}
$$

(41)

for $t \geq 1$ and given some initial value $P_{0\mid 0}$. At first, the time subscript on $\Omega_{t,P}$ and $G_t$ suggests that there remains a gap between this process and the conditional mean representation in this paper. However, standard results show that $P_{t\mid t}$ converges quickly to

$$
P = F[P - PH(H'PH + R)^{-1}H'P]F' + Q
$$

if a solution exists. (This is an Algebraic Ricatti Equation.) In addition, filtering $\hat{P}_t$ from observations of $\mathcal{P}_t$ is a special case where:

$$
H = 1 \quad F = \phi_w \quad Q = \Omega_w.
$$

Then, the steady-state matrix $P$, $G$ and $\Omega_P$ are given by

$$
\begin{align*}
P &= \phi_w[P - P(P + R)^{-1}P]\phi_w' + \Omega_w \\
\Omega_P &= P + R \\
G &= P(\Omega_P)^{-1},
\end{align*}
$$

(42)

and the steady-state process of the observed portfolios $\mathcal{P}_t$ has the following conditional mean representation:

$$
\begin{align*}
\mathcal{P}_{t+1} &= \mathcal{E}_{\mathcal{P},t} + \eta_{t+1} \\
\mathcal{E}_{\mathcal{P},t+1} &= \mu_w + \phi_w \mathcal{E}_{\mathcal{P},t} + \phi_w P(P + R)^{-1} \eta_{t+1}
\end{align*}
$$

(43)

with $\mu_w$ and $\phi_w$ given in Equation 37, $P$ and $R$ are given in Equation 42, and where $\eta_{t+1} \sim N(0, P + R)$. As expected, the Markovian case is nested with $R = 0$ since then we have that $P(P + R)^{-1} = I$ and that $P = \Omega$. 

27
A.2 First-Order Approximation

This section derives a first-order approximation to the steady-state matrices in Equation 42 around the point $R = 0$. We use the standard Kalman filter notation for the state-space system:

\begin{align*}
X_{t+1} &= FX_t + V_{t+1} \\
Y_t &= H'X_t + W_t,
\end{align*}

(44) (45)

where $X_t$ is a vector of latent state variables and $Y_t$ is a vector of measurements where we have that $E[V_t] = E[W_t] = 0$, $E[V_t V_t'] = 1_{t=\tau}Q$ and $E[W_t W_t'] = 1_{t=\tau}R$. The Kalman filter estimate of $X_{t+1}$ based on the history of observations $Y_t, \ldots, Y_1$ is given by:

\begin{align*}
E_t[X_{t+1}] &= FE_{t-1}[X_t] + \psi_{t-1} (Y_t - E_{t-1}[Y_t]) \\
&= \theta_{t-1} E_{t-1}[X_t] + \psi_{t-1} Y_t
\end{align*}

(46) (47)

with $\theta_{t-1} \equiv F - \psi_{t-1} H'$ and $\psi_{t-1} \equiv FP_{t|t-1} H (H'P_{t|t-1} H + R)^{-1}$. We have that $P_{t|t-1} \rightarrow P$ as $t$ increases where $P$ is the solution to

\[ P = F[P - PH(H'PH + R)^{-1}H'P]F' + Q \]

and, therefore, $\theta_t$ converges to $\theta = F - KH'$ with $K = FP(H'PH + R)^{-1}$. How far is $\theta_t$ from $0$? Note that $\theta_t = 0$ if $R = 0$. We derive a first-order approximation of $\theta$ around $R = 0$. We have

\[ d\theta = -(dK)H', \]

(48)

where

\begin{align*}
dK &= d(FPH(H'PH + R)^{-1}) \\
&= d(FPH)(H'PH + R)^{-1} + (FPH)d[(H'PH + R)^{-1}] \\
&= (Fd(P)H)(H'PH + R)^{-1} - (FPH)(H'PH + R)^{-1} d[H'PH + R][H'PH + R]^{-1} \\
&= (Fd(P)H)(H'PH + R)^{-1} - K[H'd(P)H + dR](H'PH + R)^{-1},
\end{align*}

and therefore,

\begin{align*}
d(\text{vec}(K)) &= \text{vec}(d(K)) \\
&= \text{vec}[(Fd(P)H)(H'PH + R)^{-1}] - \text{vec}[K[H'd(P)H + dR](H'PH + R)^{-1}] \\
&= ((H'PH + R)^{-1}H') \otimes F \text{vec}(d(P)) \\
&- (K \otimes (KH')) \text{vec}(d(P)) \\
&- (H'PH + R)^{-1} \otimes K) \text{vec}(d(R)) \\
&= ((H'PH + R)^{-1}H') \otimes \theta \text{vec}(d(P)) \\
&- (H'PH + R)^{-1} \otimes K) \text{vec}(d(R)).
\end{align*}
Consider $\text{vec}(d(P))$:

$$d(P) = Fd(P)F' - d(K)(H'PF') - KH'd(P)F'$$

$$\text{vec}(d(P)) = (F \otimes F)\text{vec}(d(P)) - ((FPH) \otimes I_r)\text{vec}(d(K)) - (F \otimes (KH'))\text{vec}(d(P))$$

$$= (F \otimes \theta)\text{vec}(d(P))$$

$$-((FPH) \otimes I_r) \left[(((H'PH + R)^{-1}H') \otimes \theta)\text{vec}(d(P)) - ((H'PH + R)^{-1} \otimes K)d(\text{vec}(R))\right]$$

$$= (\theta \otimes \theta)\text{vec}(d(P)) + (K \otimes K)d(\text{vec}(R)),$$

then a first-order approximation is given by:

$$\text{vec}(P) = (F \otimes F)\text{vec}(R) \quad \text{or} \quad P = FRF'$$

implying that:

$$d(\text{vec}(K)) = \left\{\left(((H'PH + R)^{-1}H') \otimes \theta\right[\left((I_r - \theta \otimes \theta)^{-1}(K \otimes K) - (H'PH + R)^{-1} \otimes K\right)\right\}d(\text{vec}(R))$$

The derivatives in the case $R = 0$:

$$\frac{d(\text{vec}(\theta))}{d(\text{vec}(R))}|_{R = 0} = (H \otimes I_r)[(H'P(0)H)^{-1} \otimes K(0)]$$

$$= (H(H'QH)^{-1}) \otimes (FQH(H'QH)^{-1}),$$

and a Taylor approximation of $\theta$ is given by

$$\text{vec}(\theta) \approx \left(\frac{d(\text{vec}(\theta))}{d(\text{vec}(R))}|_{R = 0}\right)\text{vec}(R)$$

$$= (H(H'QH)^{-1}) \otimes (FQH(H'QH)^{-1})\text{vec}(R),$$

or

$$\theta \approx FQH(H'QH)^{-1}R(H'QH)^{-1}H'.$$ (49)

In the case $H = I$, we have

$$\text{vec}(\theta) \approx (Q^{-1} \otimes F)\text{vec}(R),$$

or

$$\theta \approx FRQ^{-1},$$ (50)

which corresponds to Equation 32 in the text.
Figure 1: Predictability Coefficients
Coefficients on forward rates from predictability regressions of bond excess returns on a constant and the forward rates $f_t = [f_t^{(1)} \ f_t^{(2)} \ f_t^{(3)} \ f_t^{(4)} \ f_t^{(5)}]$ from a direct OLS regression, or implied from the VAR5 and the CM5 model, respectively. Each panel displays the forward rate coefficients for a given returns horizons and for bond with two years and five years to maturity. Estimation based on a sample from January 1964 until December 2007. Returns computed from the GSW dataset.

(A) 3-month

(B) 6-month

(C) 12-month
Figure 2: Model-Implied Risk Premium and the Cochrane-Piazzesi Factors
Ratios of predictive $R^2$s from regressions of the returns forecasting factor – estimated from Equation 3 – on the model-implied bond risk premiums from the $VAR_5$ relative to the $CM_5$ model. Estimation based on a sample from January 1964 until December 2007. Returns computed from the GSW dataset.

Figure 3: Yield Decomposition – 10-Year Yield
Comparison of the expectation component of the 10-year yield in the $CM_5$ and the $VAR_5$ models, respectively. Estimation based on a sample from January 1964 until December 2011.
Table 1: **Bond Excess Returns – OLS Predictability Regressions**

Predictability of excess returns on bonds with 2, 3, 4, 5, 7 and 10 years to maturity for holding horizons of 1, 2, 3, 6, 9, 12 and 24 months using annual forward rates with 1, 2, 3, 4, and 5 years to maturity. Panel (A) displays $R^2$’s predictability results from unrestricted regressions of future bond excess returns on current forward rates,

$$x_{t,h}^{(n)} = \beta_{n,h} f_t + u_{t,h},$$

where $x_{t,h}^{(n)}$ is the excess returns from holding a bond with maturity $n$ over an $h$-months horizon, $f_t$ stacks a constant with the forward rates and $\beta_{n,h}$ is a vector of coefficients. Panel (A) displays $R^2$’s predictability results from restricted regressions,

$$x_{t,h}^{(n)} = b_{n,h} \gamma_h f_t + u_{t,h},$$

where $b_{n,h}$ is a maturity-specific scalar and $\gamma_h$ is an horizon-specific vector of coefficients. Returns computed monthly from the GSW dataset between Jan. 1963 and Dec. 2003.

### Panel (A) Unrestricted OLS Regressions

<table>
<thead>
<tr>
<th>Maturity</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>6</th>
<th>9</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3,1</td>
<td>6,8</td>
<td>10,3</td>
<td>20,4</td>
<td>30,0</td>
<td>34,8</td>
</tr>
<tr>
<td>3</td>
<td>3,1</td>
<td>6,9</td>
<td>10,4</td>
<td>20,9</td>
<td>30,5</td>
<td>35,2</td>
</tr>
<tr>
<td>4</td>
<td>3,2</td>
<td>7,0</td>
<td>10,4</td>
<td>21,2</td>
<td>31,1</td>
<td>36,0</td>
</tr>
<tr>
<td>5</td>
<td>3,2</td>
<td>6,9</td>
<td>10,3</td>
<td>21,2</td>
<td>31,4</td>
<td>36,6</td>
</tr>
<tr>
<td>7</td>
<td>3,1</td>
<td>6,7</td>
<td>9,7</td>
<td>20,6</td>
<td>31,4</td>
<td>37,2</td>
</tr>
<tr>
<td>10</td>
<td>2,9</td>
<td>6,0</td>
<td>8,6</td>
<td>19,2</td>
<td>30,3</td>
<td>37,1</td>
</tr>
</tbody>
</table>

### Panel (B) Restricted OLS Regressions

<table>
<thead>
<tr>
<th>Maturity</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>6</th>
<th>9</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2,9</td>
<td>6,4</td>
<td>9,8</td>
<td>19,5</td>
<td>28,5</td>
<td>32,5</td>
</tr>
<tr>
<td>3</td>
<td>3,0</td>
<td>6,8</td>
<td>10,2</td>
<td>20,6</td>
<td>30,0</td>
<td>34,4</td>
</tr>
<tr>
<td>4</td>
<td>3,1</td>
<td>6,9</td>
<td>10,3</td>
<td>21,1</td>
<td>31,0</td>
<td>35,8</td>
</tr>
<tr>
<td>5</td>
<td>3,2</td>
<td>6,9</td>
<td>10,3</td>
<td>21,2</td>
<td>31,4</td>
<td>36,6</td>
</tr>
<tr>
<td>7</td>
<td>3,1</td>
<td>6,6</td>
<td>9,7</td>
<td>20,5</td>
<td>31,3</td>
<td>37,1</td>
</tr>
<tr>
<td>10</td>
<td>2,7</td>
<td>5,9</td>
<td>8,4</td>
<td>18,9</td>
<td>30,0</td>
<td>36,7</td>
</tr>
</tbody>
</table>
Table 2: **Bond Excess Returns – Recursive Predictability Regressions**

Predictability of excess returns on bonds with 2, 3, 4, 5, 7 and 10 years to maturity for holding horizons of 1, 2, 3, 6, 9, 12 and 24 months using annual forward rates with 1, 2, 3, 4, and 5 years to maturity. Excess returns forecasts are are derived from the non-markovian equation,

\[
x_{t,h}^{(n)} = (1 - \alpha_{n,h})\gamma'_{n,h}f_t + \alpha_{n,h}R_{t-1,h}^{(n)} + u_{t,h}
\]

\[
R_{t,h}^{(n)} = \alpha_{n,h}R_{t-1,h}^{(n)} + (1 - \alpha_{n,h})\gamma'_{n,h}f_t,
\]

where \(x_{t,h}^{(n)}\) is the excess returns from holding a bond with maturity \(n\) over an \(h\)-months horizon, \(f_t\) stacks a constant with the forward rates, \(\alpha_{n,h}\) is a scalar and \(\gamma_{n,h}\) is a vector of coefficients. Panel (A) displays \(R^2\)'s and Panel (B) displays estimates of \(\alpha_{n,h}\). Returns computed monthly from the GSW dataset between Jan. 1963 and Dec. 2003.

### Panel (A) \(R^2\)'s

<table>
<thead>
<tr>
<th>Maturity</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>6</th>
<th>9</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6.3</td>
<td>9.8</td>
<td>14.2</td>
<td>25.6</td>
<td>34.6</td>
<td>34.2</td>
</tr>
<tr>
<td>3</td>
<td>6.6</td>
<td>10.6</td>
<td>15.8</td>
<td>28.5</td>
<td>37.6</td>
<td>37.9</td>
</tr>
<tr>
<td>4</td>
<td>6.9</td>
<td>11.3</td>
<td>17.0</td>
<td>31.1</td>
<td>40.3</td>
<td>41.0</td>
</tr>
<tr>
<td>5</td>
<td>7.0</td>
<td>11.7</td>
<td>17.9</td>
<td>33.1</td>
<td>42.3</td>
<td>43.2</td>
</tr>
<tr>
<td>7</td>
<td>6.9</td>
<td>11.9</td>
<td>18.7</td>
<td>35.5</td>
<td>44.9</td>
<td>46.2</td>
</tr>
<tr>
<td>10</td>
<td>5.3</td>
<td>11.6</td>
<td>18.5</td>
<td>36.8</td>
<td>46.8</td>
<td>48.7</td>
</tr>
</tbody>
</table>

### Panel (B) \(\alpha_{n,h}\)'s

<table>
<thead>
<tr>
<th>Maturity</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>6</th>
<th>9</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.81</td>
<td>0.80</td>
<td>0.79</td>
<td>0.75</td>
<td>0.67</td>
<td>0.55</td>
</tr>
<tr>
<td>3</td>
<td>0.81</td>
<td>0.81</td>
<td>0.80</td>
<td>0.76</td>
<td>0.68</td>
<td>0.58</td>
</tr>
<tr>
<td>4</td>
<td>0.82</td>
<td>0.82</td>
<td>0.80</td>
<td>0.76</td>
<td>0.69</td>
<td>0.60</td>
</tr>
<tr>
<td>5</td>
<td>0.83</td>
<td>0.82</td>
<td>0.81</td>
<td>0.77</td>
<td>0.70</td>
<td>0.62</td>
</tr>
<tr>
<td>7</td>
<td>0.84</td>
<td>0.84</td>
<td>0.82</td>
<td>0.78</td>
<td>0.72</td>
<td>0.64</td>
</tr>
<tr>
<td>10</td>
<td>0.16</td>
<td>0.85</td>
<td>0.83</td>
<td>0.79</td>
<td>0.73</td>
<td>0.66</td>
</tr>
</tbody>
</table>
Table 3: Bond Excess Returns – Recursive Predictability Regressions
Predictability of excess returns on bonds with 2, 3, 4, 5, 7 and 10 years to maturity for holding horizons of 1, 2, 3, 6, 9, 12 and 24 months using annual forward rates with 1, 2, 3, 4, and 5 years to maturity. Excess returns forecasts are derived from the non-markovian equation,

\[ x_{t,h}^{(n)} = b_{n,h} \left( (1 - \alpha_h) \gamma_h f_t + \alpha_h R_{t-1,h} \right) + u_{t,h}^{(n)} \]

\[ R_{t,h} = \alpha_h R_{t-1,h} + (1 - \alpha_h) \gamma_h f_t, \]  

where \( x_{t,h}^{(n)} \) is the excess returns from holding a bond with maturity \( n \) over an \( h \)-months horizon, \( f_t \) stacks a constant with the forward rates, \( \beta_{n,h} \) is a scalar, \( \alpha_h \) is a scalar and \( \gamma_h \) is a vector of coefficients. Panel (A) displays \( R^2 \)’s, Panel (B) displays estimates of \( \alpha_h \), Panel (C) displays estimates of \( b_{n,h} \) and Panel (D) displays estimates of \( \gamma_h \). Returns computed monthly from the GSW dataset between Jan. 1963 and Dec. 2003.

Panel (A) \( R^2 \)’s

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Holding Period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>6.6</td>
</tr>
<tr>
<td>3</td>
<td>7.0</td>
</tr>
<tr>
<td>4</td>
<td>7.2</td>
</tr>
<tr>
<td>5</td>
<td>7.3</td>
</tr>
<tr>
<td>7</td>
<td>7.1</td>
</tr>
<tr>
<td>10</td>
<td>6.5</td>
</tr>
</tbody>
</table>

Panel (B) \( \alpha_h \)’s

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Holding Period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.8320</td>
</tr>
</tbody>
</table>

Panel (C) \( b_{n,h} \)’s

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Holding Period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.47</td>
</tr>
<tr>
<td>3</td>
<td>0.67</td>
</tr>
<tr>
<td>4</td>
<td>0.85</td>
</tr>
<tr>
<td>5</td>
<td>1.01</td>
</tr>
<tr>
<td>7</td>
<td>1.30</td>
</tr>
<tr>
<td>10</td>
<td>1.69</td>
</tr>
</tbody>
</table>

Panel (D) \( \gamma_h \)’s

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Holding Period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>-0.04</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>-9.53</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>8.87</td>
</tr>
<tr>
<td>( \gamma_3 )</td>
<td>11.84</td>
</tr>
<tr>
<td>( \gamma_4 )</td>
<td>2.87</td>
</tr>
</tbody>
</table>
Table 4: **Bond Excess Returns – Common Factor**
Principal component analysis of the return-forecasting factors $R_{t,h}$ estimated across horizons of 1, 2, 3, 6, 9, and 12 months (see Table 3).

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.46</td>
<td>-0.42</td>
<td>-0.60</td>
<td>-0.08</td>
<td>0.28</td>
<td>-0.41</td>
</tr>
<tr>
<td>2</td>
<td>0.45</td>
<td>-0.36</td>
<td>-0.07</td>
<td>0.23</td>
<td>-0.34</td>
<td>0.71</td>
</tr>
<tr>
<td>3</td>
<td>0.45</td>
<td>-0.16</td>
<td>0.66</td>
<td>0.24</td>
<td>-0.20</td>
<td>-0.48</td>
</tr>
<tr>
<td>6</td>
<td>0.41</td>
<td>0.17</td>
<td>0.34</td>
<td>-0.42</td>
<td>0.64</td>
<td>0.31</td>
</tr>
<tr>
<td>9</td>
<td>0.36</td>
<td>0.47</td>
<td>-0.18</td>
<td>-0.54</td>
<td>-0.57</td>
<td>-0.10</td>
</tr>
<tr>
<td>12</td>
<td>0.30</td>
<td>0.65</td>
<td>-0.22</td>
<td>0.64</td>
<td>0.18</td>
<td>-0.004</td>
</tr>
</tbody>
</table>

$R^2$ 96.8 2.7 0.31 0.11 0.02 0.003

Table 5: **Bond Excess Returns – Model-Implied Predictability**
Predictability of excess returns on bonds with 2, 3, 4, and 5 years to maturity for holding horizons of one year using annual forward rates with 1, 2, 3, 4, and 5 years to maturity implied by different time-series model on yields estimated at the monthly frequency. Stacking yields with maturities of 1, 2, 3, 4 and 5 years to maturity in the vector $y_t$, Panel A display the predictability $R^2$s implied by a VAR(1), a VARMA(1,1) and a VAR(12) model, respectively, in the case of the unrestricted OLS regression (Equation 1). Panel B displays the RMSEs from an out-of-sample exercise where the parameters are fixed at values estimated in the first half of the sample and the forecasts’ accuracy are compared using the second half of the sample. Estimation based on data between Jan. 1963 and Dec. 2003.

Panel (A) Benchmark CP Regressions

<table>
<thead>
<tr>
<th>Maturity</th>
<th>VAR1</th>
<th>VARMA11</th>
<th>VAR12</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>14.7</td>
<td>25.8</td>
<td>30.8</td>
</tr>
<tr>
<td>3</td>
<td>13.5</td>
<td>28.1</td>
<td>33.0</td>
</tr>
<tr>
<td>4</td>
<td>15.1</td>
<td>32.8</td>
<td>35.7</td>
</tr>
<tr>
<td>5</td>
<td>14.6</td>
<td>32.4</td>
<td>33.1</td>
</tr>
</tbody>
</table>

Panel (B) Out-of-Sample

<table>
<thead>
<tr>
<th>Maturity</th>
<th>VAR1</th>
<th>VARMA11</th>
<th>VAR12</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.51</td>
<td>1.31</td>
<td>1.54</td>
</tr>
<tr>
<td>3</td>
<td>2.91</td>
<td>2.46</td>
<td>2.90</td>
</tr>
<tr>
<td>4</td>
<td>4.08</td>
<td>3.37</td>
<td>3.96</td>
</tr>
<tr>
<td>5</td>
<td>5.15</td>
<td>4.25</td>
<td>4.88</td>
</tr>
</tbody>
</table>
Table 6: **Sample Bond Risk Premium**

For each of the $VAR_5$ and the $CM_5$ model, we compute the conditional bond risk premium in Equation 24 for bond maturities of 6, 12, 24 and 60 months and holding periods between 1 and 12 months at each date in the sample, and report the ratio of the sample variance in the $VAR_5$ relative to the sample variance in the $CM_5$ in each case. Estimation based on a CRSP data from January 1964 until December 2007 and returns computed from the GSW dataset.

<table>
<thead>
<tr>
<th>Panel (A) Variance ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel (B) $R^2$'s ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel (C) $R^2$'s ratios – Restricted case</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>60</td>
</tr>
</tbody>
</table>

Table 7: **Sample Bond Risk Premium - Kalman Filters**

For each version of the $CM_5 - KF$ models, the first line reports the number of additional parameters relative to the closest version (i.e., $CM_5 - KF3$ vs. $VAR_5$, $CM_5 - KF3$ vs. $CM_5 - KF2$, ...). The second line reports the corresponding likelihood gain. The remaining lines report variance ratios. For each $CM_5 - KF$ model we compute the conditional bond risk premium in Equation 24 for bond maturities of 24, 36, 48 and 60 months and for a holding period of 12 months, at each date in the sample, and report the ratio of the sample variance in the $VAR_5$ relative to the sample variance in different $CM_5 - KF$ models. Estimation based on CRSP data from January 1964 until December 2007, returns computed from the CRSP data.

<table>
<thead>
<tr>
<th>Panel (A) Variance ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Param. Incr.</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Lik. Gain</td>
</tr>
<tr>
<td>1 year</td>
</tr>
<tr>
<td>2 years</td>
</tr>
<tr>
<td>3 years</td>
</tr>
<tr>
<td>4 years</td>
</tr>
<tr>
<td>5 years</td>
</tr>
</tbody>
</table>