A Parsimonious No-Arbitrage Term Structure Model that is Useful for Forecasting *

Fan Dora Xia †

UC San Diego

First draft: August 8, 2012
This draft: November 9, 2013

Abstract

I propose a parsimonious Gaussian Affine Term Structure Model (GATSM) to reconcile empirical findings that while the level, slope and curvature (or the first three principal components of yields) can quite accurately describe the cross-section of yields, different linear combinations of interest rates and other macro variables are useful to predict excess returns. I introduce a forecasting factor, which compactly summarizes rich information in expected excess returns, to a conventional three-factor (the level, slope and curvature) GATSM. This fourth factor is constructed by reduced rank forecasting regression with a large predictor set, and it can explain one-year excess returns of two- to five-year maturity bonds from 1964 to 2007 with $R^2$ up to 0.71. Considering the fact that the forecasting factor and the first three principal components span the cross-section of expected excess returns and that of yields, respectively, I restrict parameters of the four-factor GATSM. In contrast with the conventional three-factor GATSM, the restricted four-factor GATSM generates plausible countercyclical term premia.

Keywords: Gaussian affine term structure model, macroeconomy and risk premia

JEL classification codes: E43, E44

*I would like to thank my advisor James D. Hamilton for his invaluable advice.
†Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA, faxia@ucsd.edu.
1 Introduction

Back in the 1990s, [Litterman and Scheinkman(1991)] found that more than 99% of the movements of bond yields with various maturities are captured by the first three principal components of yields. These are commonly described as level, slope, and curvature based on how shocks to these components affect the yield curve. Quite interestingly, different linear combinations of interest rates and other macro variables are revealed to be useful when the goal is to predict excess returns. [Cochrane and Piazzesi(2005)] documented a tent-shaped linear combination of forward rates that predicts excess bond returns. [Cooper and Priestly(2009)] found that the output gap is helpful for forecasting excess returns in stock and bond markets. [Ludvigson and Ng(2011)] demonstrated that factors constructed from a large panel of macroeconomic variables have significant predictive power for excess bond returns. [Dahlquist and Hasseltoft(2012)] identified local and global factors that jointly predict returns in international bond markets.

Factor structure exhibited in the cross-section of yields has been explored by many term structure models. An extensively employed approach is the Gaussian Affine Term Structure Model (GATSM). It not only condenses the rich term structure of interest rates with factors of lower dimension but also restricts the mapping between factors and yields to preclude arbitrage opportunities. Among many other applications, the model has been used to forecast yields, characterize time variations of risk premia, and assess monetary policies’ effect on the yield curve. In most of these studies, researchers implicitly assume that only factors contributing to explain the cross-section of yields are relevant. This assumption, however, contradicts the empirical evidence noted above that variables other than those summarizing the cross-section of yields predict excess returns. An apparent solution is to include all variables with predictive power as factors. But this approach would result in a large number of extra factors and make overfitting unavoidable.

[1] Piazzesi(2010), Gurkaynak and Wright(2012), and Duffee(2012) provided extensive relevant literature reviews.
In this paper I propose a parsimonious GATSM to circumvent inadequate factor specification in commonly used GATSMs without exploding the parameter space. First, I augment a conventional three-factor (the level, slope and curvature) GATSM with a forecasting factor, which compactly summarizes information in expected excess returns. The forecasting factor is constructed using reduced rank forecasting regression of excess returns on a large predictor set, including a panel of 131 macroeconomic variables and initial forward rates. The comprehensive predictor set makes use of rich information in the macroeconomy and allows relatively thorough extraction of predicted components from realized excess returns. Reduced rank forecasting regression serves the purpose of finding minimal sufficient linear combinations of regressors to predict excess returns with different maturities. This forecasting factor can explain one-year excess returns of two- to five-year maturity bonds from 1964 to 2007 with $R^2$ up to 0.71. Second, I restrict parameters of the four-factor GATSM to further shrink parameter space. Factors are constructed in such a way that while cross-sectional information in yields is spanned by the first three principal components, cross-sectional information in expected excess returns is spanned by the forecasting factor, thus leading to restrictions on parameters. To assess economic significance of the proposed model, I apply the model to decompose the yield curve. The restricted four-factor GATSM generates countercyclical term premium as economic theories suggest, while the conventional three-factor GATSM implies almost acyclical term premium.

Similar models have been proposed in other studies. Cochrane and Piazzesi (2009) used the tent-shaped linear combination of forward rates from Cochrane and Piazzesi (2005) as a fourth factor in addition to the first three principal components. They put a restriction that only level risk is priced by the fourth factor. Joslin, Priebsch, and Singleton (2010) studied a restricted GATSM with the national activity index, the inflation rate along with the level, slope and curvature. The first set of their restrictions, which is similar to a subset of restrictions derived in this paper, is imposed so that two macro factors do not help explaining contemporaneous yields conditional on the first three principal components.
Other restrictions are selected out of $2^{19}$ possibilities by model selection technique. Compared with previous research, this paper is different in the following two aspects. First, I use a considerably larger predictor set to extract more predictable components from realized excess returns. Second, restrictions are derived from factors’ information-spanning properties instead of previous experience or computationally expensive model selection procedures.

Alternative approaches without specifying factors as observables are taken by Kim and Wright (2005) and Duffee (2011). Kim and Wright (2005) specified a GATSM with three latent factors to explain both yields and survey data on expected interest rates. Duffee (2011) estimated a restricted GATSM with five latent factors, in which two additional factors allows both information in the cross-section of yields and that in the dynamics of yields to be filtered out. Compensations for risks are restricted so that price for level risk is time-varying, price for slope risk is constant and other risks are not priced. As Duffee (2011) pointed out, latent factor approach avoids the risk of misspecifying the relation between the yield curve and the macroeconomy, while models using other macroeconomic variables have more precise estimates.

The rest of the paper is organized as follows. Section 2 constructs the forecasting factor by reduced rank forecasting regression. Section 3 derives restrictions. I estimate the model in Section 4 and examine the model’s implications for the term premium in Section 5. Section 6 concludes.

2 Forecasting factor construction

2.1 GATSM

Before showing how to construct the forecasting factor, I review the GATSM briefly.
**Notation** I denote the log price of an \( n \)-period zero-coupon bond at time \( t \) by \( p_t^{(n)} \). The corresponding yield is

\[
y_t^{(n)} = -\frac{1}{n} p_t^{(n)}.
\]

The \( m \)-period forward rate between time \( t + n \) and \( t + n + m \) is defined as

\[
f_t^{(n,m)} = \frac{1}{m} \left( (n + m) y_t^{(n+m)} - n y_t^{(n)} \right).
\]

The holding period return on buying an \( n \)-period zero-coupon bond at time \( t \) and then selling it as an \( (n - m) \)-period zero-coupon bond at time \( t + m \) is given by

\[
hpr_t^{(n)} = \frac{1}{m} \left( n y_t^{(n)} - (n - m) y_{t+m}^{(n-m)} \right).
\]

The difference between \( hpr_t^{(n)} \) and \( y_t^{(m)} \) is the \( m \)-period excess return of an \( n \)-period bond:

\[
exr_t^{(n)} = \frac{1}{m} \left( n y_t^{(n)} - (n - m) y_{t+m}^{(n-m)} - m y_{t}^{(m)} \right).
\]

**Model specification** The GATSM assumes that there are \( K \) factors, denoted by \( X_t \), relevant for bond pricing. These factors follow a first order vector autoregressive process (VAR(1)) under the physical (\( P \)) measure:

\[
X_{t+1} = \mu + \rho X_t + \Sigma \varepsilon_{t+1}, \varepsilon_{t+1} \sim N(0, I).
\]

Log stochastic discount factor is essentially affine as in [Duffee(2002)]

\[
m_{t+1} = -r_t - \frac{1}{2} \lambda_t \lambda_t - \lambda_t \varepsilon_{t+1},
\]
where \( r_t \) is the short rate, and \( \lambda_t \) is the price of risk which is linear in the factors, i.e.

\[
\lambda_t = \lambda_0 + \lambda_1 X_t.
\]  

(3)

With assumptions (1)-(3), it can be shown that any assets with payoff being a function of the factors \( g(X_{t+1}) \) can be priced in the following way (see Appendix 7 for the derivation):

\[
\text{Price}(X_t) = \exp(-r_t) \mathbb{E}_t^Q(g(X_{t+1})).
\]

The expectation is taken under the risk neutral \((Q)\) measure, in which \( X_t \) follows a VAR(1) as well:

\[
X_{t+1} = \mu^Q + \rho^Q X_t + \Sigma^Q \varepsilon_{t+1}^Q \quad \text{where} \quad \varepsilon_{t+1}^Q \sim N(0, I),
\]

with

\[
\mu^Q = \mu - \Sigma \lambda_0 \quad \text{and} \quad \rho^Q = \rho - \Sigma \lambda_1.
\]

If we further assume that the short rate is an affine function of the factors, i.e.

\[
y_t^{(1)} = r_t = \delta_0 + \delta_1' X_t,
\]  

(4)

then we can derive analytical expressions for yields of zero-coupon bonds with all maturities. It turns out that they are affine functions of the factors as well (see Appendix 7 for the derivation), i.e.

\[
y_t^{(n)} = \frac{a_n}{n} + \frac{b_n}{n} X_t,
\]  

(5)
where \( a_n \) and \( b_n \) can be calculated recursively for \( n \geq 2 \) as follows:

\[
\begin{align*}
    a_n &= a_{n-1} + \delta_0 + b'_{n-1} \mu^Q - \frac{1}{2} b'_{n-1} \Sigma \Sigma' b_{n-1}, \\
    b'_n &= b'_{n-1} \rho^Q + \delta'_1.
\end{align*}
\]

(6)

The recursion starts with initial conditions: \( a_1 = \delta_0 \) and \( b'_1 = \delta'_1 \).

2.2 Reduced rank forecasting regression

In the GATSM, \( X_t \) determines both yields and expected excess returns. When taking the model to data, most researchers focus solely on variables that explain the cross-section of yields as if only they are relevant. This common practice leaves out factors that are not spanned by the cross-section of yields but can predict excess returns. To correct model misspecification mentioned above, we need to augment conventional GATSMs with factors conveying information in expected excess returns. At the same time, we do not want to increase the number of factors substantially due to overfitting concern. Therefore, it is desirable to use factors of low dimension to summarize the cross-section of expected excess returns in a similar way the level, slope and curvature achieve for the cross-section of yields. Principal component analysis, through which the level, slope and curvature are obtained, can not be applied since only realized excess returns are observed. For unobserved expected excess returns, fitted values from forecasting regressions are often used as proxies. Most researchers focus on a handful of predictor variables to limit parameter space. But this approach increases the chance of missing predictable components in realized returns. To address this concern, [Ludvigson and Ng(2011)] introduced a comprehensive panel of 131 macroeconomic variables and used dynamic factor analysis to mitigate overparameterization. In their paper, they first constructed eight factors from the panel by principal component analysis or Gibbs sampling. Then they chose macroeconomic predictors from these eight factors, their square and cubic terms by an extensive model selection process. Note that they selected different
sets of predictors for excess returns with different maturities. Though replacing 131 macroeconomic variables with eight factors greatly reduces the dimension of predictor set, these eight factors were constructed solely from macroeconomic series to describe the associations among them rather than to predict excess returns. Also, studying excess returns individually does not fulfill the aim of consolidating cross-sectional information in expected excess returns. In order to extract more predictable components from realized returns, I include the large panel of macroeconomic variables in my predictor set as well. Different from Ludvigson and Ng(2011), I employ reduced rank regression to shrink the parameter space. Reduced rank regression uses low dimension linear combinations of regressors to model the variation in regressands by making use of the fact that regressands are likely to be correlated.

Regression specification  The reduced rank forecasting regression is specified as follows

$$
\overline{exr}_{t,t+12} = c_r + \alpha_r \beta_r \left( Z_t' \ F_t' \right)' + e_{r,t}.
$$

The dependent variable $\overline{exr}_{t,t+12}$ is the $4 \times 1$ column vector of standardized one-year excess returns of two- to five-year maturity bonds, i.e.

$$
\overline{exr}_{t,t+12} = \left( \overline{exr}_{t,t+12}^{(24)} \ \ \overline{exr}_{t,t+12}^{(36)} \ \ \overline{exr}_{t,t+12}^{(48)} \ \ \overline{exr}_{t,t+12}^{(60)} \right)'.
$$

For the predictors, $Z_t$ collects 131 macroeconomic variables, and $F_t$ includes the one-year forward rates of one- to five-year maturity bonds, i.e.

$$
F_t = \left( f_t^{(0,12)} \ \ f_t^{(12,12)} \ \ f_t^{(24,12)} \ \ f_t^{(36,12)} \ \ f_t^{(48,12)} \right)'.
$$

The matrices $\alpha_r$ and $\beta_r$ are of dimension $4 \times r$ and $r \times 136$, respectively, where $r$ is the rank to be specified. By using the product of $\alpha_r$ and $\beta_r$ as the coefficient, equation 7 assumes

Anderson(1951) first introduced reduced rank regression, and Reinsel and Velu(1998) offered a comprehensive review.
that the variation of excess returns is predicted by $r$ linear combinations of regressors, i.e.

$$\beta_r \left( Z_t' \ F_t' \right)' .$$

Compared with unrestricted regressions, reduced rank regression shrinks the number of parameters by $(4 - r)(136 - r)^3$ The constant term and the forecasting error are represented by $c_r$ and $e_{r,t}$, respectively. Both $c_r$ and $e_{r,t}$ are $4 \times 1$ column vectors. I specify $e_{r,t} \sim N(0, \sigma^2_r I)$ with two considerations. First, as pointed out in Reinsel and Velu(1998), when the number of observations is not large relative to the dimensions of regressors and regressands, estimates allowing general error covariance matrix may not be very accurate and specifying $\text{Cov}(e_{r,t}) = \sigma^2_r I$ with standardized response variables is preferred.\(^4\) Also, my purpose here is to extract factors to explain the total variance of excess returns instead of capturing their covariance. I thus impose diagonal error covariance matrix. With returns being brought to same scale, I further assume that diagonal elements are identical.

For estimation purpose, $\alpha_r$, $\beta_r$, $c_r$ and $\sigma^2_r$ are chosen to maximize the log likelihood. Corresponding expressions can be found in Reinsel and Velu(1998) and are given by

\[
\hat{\alpha}_r = \begin{pmatrix} \hat{V}_1' & \cdots & \hat{V}_r' \end{pmatrix},
\]

\[
\hat{\beta}_r = \begin{pmatrix} \hat{V}_1' \\ \vdots \\ \hat{V}_r' \end{pmatrix} \hat{\Sigma}_{z_r z_f}^{-1} \hat{\Sigma}_{z_f z_f} \hat{\Sigma}_{z_r z_f}',
\]

\[
\hat{c}_r = \frac{1}{T} \sum_{t=1}^{T} \left( \hat{e}_{F t,t+12} - \frac{1}{T} \sum_{t=1}^{T} \hat{\alpha}_r \hat{\beta}_r \left( Z_t' \ F_t' \right)' \right),
\]

\[
\hat{\sigma}^2_r = \frac{1}{T \times 4} \sum_{t=1}^{T} \left( \hat{e}_{F t,t+12} - \hat{c}_r - \hat{\alpha}_r \hat{\beta}_r \left( Z_t' \ F_t' \right)' \right) \left( \hat{e}_{F t,t+12} - \hat{c}_r - \hat{\alpha}_r \hat{\beta}_r \left( Z_t' \ F_t' \right)' \right)'.
\]

In the above expressions, $T$ is the number of observations, $\hat{\Sigma}_{z_f z_f}$ and $\hat{\Sigma}_{z_r z_f}$ are respectively the sample variance of the predictors and the sample covariance between the predictors and the standardized excess returns, and $\hat{V}_j$ is the eigenvector of $\hat{\Sigma}_{z_r z_f} \hat{\Sigma}_{z_f z_f}^{-1} \hat{\Sigma}_{z_r z_f}'$ as-

\(^3\)Though the parameter space is still large compared with that in Ludvigson and Ng(2011) even for the lowest possible rank $r = 1$, reduced rank regression can be quite powerful in applications with a large cross-section of excess returns.

\(^4\)In my case, I have 136 predictors with only 528 observations.
associated with the $j$th largest eigenvalue. The corresponding maximized log likelihood is

$$
L_r = -\frac{Tx^4}{2} \log (2\pi) - \frac{Tx^4}{2} \log (\hat{\sigma}^2) - \frac{T^4}{2}.
$$

### Data

I use the end-of-month Fama-Bliss data of one- through five-year zero coupon bond prices from January 1964 to December 2007, and compute yields, forward rates and excess returns accordingly. The Fama-Bliss data is not smoothed across maturities and has been widely used to study excess bond returns. For $Z_t$, I use the panel of 131 monthly macroeconomic variables in Ludvigson and Ng (2011). The panel represents eight broad categories of macroeconomic time series, including output and income, labor market, housing, consumption, orders and inventories, money and credit, bond and exchange rates, prices and stock market. It offers a comprehensive description of the macroeconomy. All series are stationarized and standardized before regressions. The sample starts from 1964 due to the availability of bond price data and ends before the Zero Lower Bound (ZLB) period, during which the GATSM is not appropriate to apply as Wu and Xia (2013) pointed out.

### Rank selection

How many factors we use to model the variation in expected returns depends on our choice of rank $r$. I let the data decide using likelihood ratio tests. First I test $H_0 : \text{rank} = r$ over alternative full rank hypothesis. I compute $-2 (L_r - L_4)$ for $r = 1$, 2, and 3, and I reject $H_0$ if the test statistic is greater than the critical value determined by the $\chi^2_{(4-r)(136-r)}$ distribution. The smallest value of $r$ for which $H_0$ is not rejected provides a good candidate. Respective test statistics for $r = 1, 2, \text{and } 3$ are 76.86, 13.27, and 4.14. Corresponding critical values with 95% level of confidence are 452.92, 307.18, and 160.91, respectively. Thus this test proposes $r = 1$. Another possible guidance can be obtained by considering the following hypothesis test $H_0 : \text{rank} = r$ v.s. $H_A : \text{rank} = r + 1$ for $r = 1, 2, \text{and } 3$. The null hypothesis is rejected if $-2 (L_r - L_{r+1})$ is above the critical value from

---

5 The data can be obtained from the Center for Research in Security Prices (CRSP).

6 The data can be downloaded from Serena Ng’s website: [http://www.columbia.edu/~sn2294/pub.html](http://www.columbia.edu/~sn2294/pub.html)

Thanks to Sydney C. Ludvigson and Serena Ng for making their data public available.

7 A detailed description of each series and its corresponding transformation to ensure stationarity can be found in Data Appendix of Ludvigson and Ng (2011).
\( \chi^2_{(4+136-(2r+1))} \). And the smallest \( r \) that \( H_0 \) is not rejected gives another choice. Respective test statistics for \( r = 1, 2, \) and \( 3 \) are 63.60, 9.12, and 4.14, while critical values with 95% level of confidence are 165.32, 163.12, and 160.91, respectively. The second test recommends \( r = 1 \) as well. Both tests suggest that all we need is one factor to capture the cross-section of expected excess returns. Compared with the most flexible forecasting regression, reduced rank regression with \( r = 1 \) decreases the number of parameters by 74%.

**Forecasting factor** With \( r = 1 \), I calculate \( \hat{\beta}_1 \) and compute \( H_t = \hat{\beta}_1 \begin{pmatrix} Z_t' & F_t' \end{pmatrix}' \). Since reduced rank forecasting regression suggests that this particular linear combination of regressors is sufficient to predict excess returns, I call \( H_t \) the forecasting factor. Before introducing the forecasting factor to a GATSM, it is important to check how well it predicts individual excess returns. I run the following OLS forecasting regressions for \( n = 24, 36, 48 \) and \( 60 \):

\[
exr_t^{(n)}_{t,t+12} = \alpha_h^{(n)} + \beta_h^{(n)} H_t + e_{h,t}^{(n)}. \tag{8}
\]

\( R^2 \)'s are 0.70, 0.71, 0.71, and 0.68, respectively. I also plot expected excess returns predicted by \( H_t \) with realized returns in Figure 1. It can be seen that predicted excess returns track the dynamics of realized returns reasonably well across all maturities. Indeed, the constructed forecasting factor well summarizes the cross-section of expected excess returns.

A well known factor that can predict excess bond returns is the tent-shaped linear combination of forward rates proposed by Cochrane and Piazzesi(2005), denoted by \( CP_t \). It is interesting to make a comparison between \( H_t \) and \( CP_t \). I construct \( CP_t \) by first running a OLS forecasting regression of the average excess return on all forward rates:

\[
\frac{1}{4} (exr_t^{(24)} + exr_t^{(36)} + exr_t^{(48)} + exr_t^{(60)}) = \gamma_0 + \begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \gamma_5 \end{pmatrix} F_t + e_{cp,t}. \tag{9}
\]

Estimated coefficients \( \hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\gamma}_4, \hat{\gamma}_5 \) are plotted in Figure 2. The tent shape remains robust.
Then $CP_t$ can be computed by $CP_t = \left( \hat{\gamma}_1 \ \hat{\gamma}_2 \ \hat{\gamma}_3 \ \hat{\gamma}_4 \ \hat{\gamma}_5 \right) F_t$. $R^2$s from regressing individual excess returns on $CP_t$ with a constant term are 0.26, 0.27, 0.30, and 0.28 for $n = 24, 36, 48$ and 60, respectively. Though $CP_t$ stays a quite useful predictor in the extended sample, $H_t$ demonstrates more in-sample predictive power with higher $R^2$s. Next I examine whether $CP_t$ remains significant conditional on $H_t$ by the following augmented forecasting regressions

$$exr_{t,t+12}^{(n)} = \alpha_{hcp}^{(n)} + \beta_{hcp}^{(n)} H_t + \gamma_{hcp}^{(n)} CP_t + \epsilon_{hcp,t}^{(n)}$$

for $n = 24, 36, 48$, and 60. Respective $R^2$s from OLS regressions specified in equation 10 are 0.70, 0.71, 0.71, and 0.68, the same with regressions without $CP_t$ included in equation 8. And $\gamma_{hcp}^{(n)} = 0$ can not be rejected at any conventional significance level for all $n$ since corresponding t-statistics with Newey-West standard errors are 0.62, 0.21, 0.56, and 0.18, respectively. The above exercise shows that $H_t$ subsumes $CP_t$’s role in in-sample prediction. If $CP_t$ instead of $H_t$ is used as the forecasting factor, certain predictable information in realized excess returns is neglected.

### 3 Restrictions derivation

Section 2 constructs the forecasting factor with which we can efficiently incorporate information in expected excess returns to a GATSM. This section aims for further model simplification by putting restrictions on parameters. It is not uncommon to restrict parameters of GATSMs. Restrictions can be set either by zeroing out insignificant parameters from estimating more flexible models or by researchers’ economic intuition on risk pricing, with Duffee(2002), Ang and Piazzesi(2003), Cochrane and Piazzesi(2009) and Duffee(2011) as examples. In this paper, I take a less ad hoc approach and impose restrictions based on factors’ information-spanning properties.

---

In this paper, I add the forecasting factor to the conventional three-factor GATSM. Therefore, \( X_t \) consists of two sets of factors, i.e. \( X_t = \left( P'_t \ H'_t \right)' \). The first set of factors \( P_t \) includes the level, slope and curvature, which are constructed from principal component analysis of yields. Together the level, slope and curvature explain 98.61%+1.33%+0.03% = 99.97% of the total variance of yields. The second set of factor \( H_t \) is the forecasting factor constructed in Section 2. It spans the cross-section of expected excess returns. The time series for these factors are plotted in Figure 3. Note that conditional on \( P_t \), \( H_t \) does not offer additional explanation power on current yields. On the other hand, conditional on \( H_t \), \( P_t \) does not help to explain expected one-year excess returns. As a result, corresponding factor loadings must be zero. And restrictions can be derived accordingly.

### 3.1 Restriction I:

\[
\delta_{1,H} = 0 \quad \text{and} \quad \rho_{P,H}^Q = 0. 
\]

(11)

This set of restrictions is based on the idea that the factor loadings of yields on the forecasting factor should be zero.

The short rate loads on the factors in the following way:

\[
yt^{(1)} = \delta_0 + \delta_1' X_t \\
= \delta_0 + \left( \delta_{1,P}' \ \delta_{1,H}' \right) \begin{pmatrix} P_t \\ H_t \end{pmatrix} \\
= \delta_0 + \delta_{1,P}' P_t + \delta_{1,H}' H_t.
\]

It is straightforward to see that the factor loading on the forecasting factor of the short rate
is zero if and only if $\delta_{1,H} = 0$. Assuming $b_{n,H} = 0$, we have

$$
\begin{align*}
    b'_{n+1} &= b'_n \rho^Q + \delta'_1 \\
    &= \left( b'_{n,P} \quad b'_{n,H} = 0 \right) \begin{pmatrix} \rho^Q_{PP} & \rho^Q_{PH} \\ \rho^Q_{HP} & \rho^Q_{HH} \end{pmatrix} + \left( \delta'_{1,P} \quad \delta'_{1,H} = 0 \right) \\
    &= \left( b'_{n,P} \rho^Q_{PP} \quad b'_{n,H} \rho^Q_{PH} \right) + \left( \delta'_{1,H} \quad 0 \right).
\end{align*}
$$

It is not hard to see that $\rho^Q_{PH} = 0$ is the necessary and sufficient condition to get the induction $b_{n,H} = 0 \Rightarrow b_{n+1,H} = 0$ going through.

Intuitively, the restriction $\rho^Q_{PH} = 0$ can be understood by the following argument. The yield of an $n$-period zero-coupon bond is the average of expected future short rates between $t$ and $t+n-1$ under the $Q$ measure plus a constant, i.e. $y_t^{(n)} = \frac{1}{n} \left( r_t + E^Q_t (r_{t+1}) + ... + E^Q_t (r_{t+n-1}) \right) + \text{constant}$. Since short rates only depend on $P_t$ (guaranteed by $\delta_{1,H} = 0$), if and only if $H_t$ does not affect time evolution of $P_t$ under the $Q$ measure, it would have no effect on yields with arbitrary maturity. The irrelevance of $H_t$ on dynamics of $P_t$ under the $Q$ measure can be achieved by setting $\rho^Q_{PH}$ to zero. Thus, only the upper left block of $\rho^Q$ is relevant for bond pricing. The upper right block is all zeros and lower blocks are not identified from observed yields, i.e. $\rho^Q = \begin{pmatrix} \rho^Q_{PP} & 0 \\ \text{not identified} & \end{pmatrix}$.

### 3.2 Restriction II:

$$
((\rho^Q)^m)_{PP} = (\rho^m)_{PP}.
$$

This restriction is based on the idea that the factor loadings on $P_t$ of expected $m$-period excess returns should be zero.\footnote{Since I focus on one-year excess returns, $m$ is fixed to be 12 for the rest of the paper.}

Expected excess returns depend on the factors and their expectations in the following
way:

\[ m \mathbb{E}_t (\text{exr}_{t,t+m}^{(n,m)}) = n y_t^{(n)} - (n - m) \mathbb{E}_t (y_{t+m}^{(n-m)}) - m y_t^{(m)} \]

\[ = a_n - a_{n-m} - a_m + (b'_n - b'_m)X_t - b'_{n-m} \mathbb{E}_t (X_{t+m}) . \]

Since

\[ \mathbb{E}_t (X_{t+m}) = (I + \rho + \ldots + \rho^{m-1}) \mu + \rho^m X_t, \]

\[ b'_n - b'_m = b'_{n-m} (\rho^Q)^m , \]

we have

\[ m \mathbb{E}_t (\text{exr}_{t,t+m}^{(n,m)}) = a_n - a_{n-m} - a_m - b'_{n-m} (I + \rho + \ldots + \rho^{m-1}) \mu \]

\[ + b'_{n-m} ((\rho^Q)^m - \rho^m) X_t . \quad (13) \]

Define \( \Lambda_m \equiv (\rho^Q)^m - \rho^m \), then

\[ b'_{n-m} \Lambda_m = \begin{pmatrix} b'_{n-m,P} & b'_{n-m,H} = 0 \end{pmatrix} \begin{pmatrix} \Lambda_{m,PP} & \Lambda_{m,PH} \\ \Lambda_{m,HP} & \Lambda_{m,PP} \end{pmatrix} \]

\[ = \begin{pmatrix} b'_{n-m,P} \Lambda_{m,PP} & b'_{n-m,P} \Lambda_{m,PH} \end{pmatrix} . \]

The factor loading of interest is \( b'_{n-m,P} \Lambda_{m,PP} \). It is a zero vector if and only if \( \Lambda_{m,PP} = 0 \).

By the definition of \( \Lambda_m \), restriction 12 can be obtained.

There is an alternative derivation. Note that

\[ n y_t^{(n)} - m y_t^{(m)} = \mathbb{E}_t^Q (r_{t+m}) + \ldots + \mathbb{E}_t^Q (r_{t+n-1}) + \text{constant} \]

\[ = (n - m) \mathbb{E}_t^Q (y_{t+m}^{(n-m)}) + \text{constant} . \]
Then the expected \( m \)-period excess return is

\[
mE_t \left( e_{xr_{t,t+m}}(n) \right) = (n - m) \left( E_t^Q - E_t \right) \left( y_{t+m}^{(n-m)} \right) + \text{constant}
\]

\[
= b'_{n-m} \left( E_t^Q - E_t \right) (X_{t+m}) + \text{constant}
\]

\[
= b'_{n-m} \left( \left( \rho^Q \right)^m - \rho^m \right) X_t + \text{constant}. \quad (14)
\]

Except for the unspecified constant part, the resulting expression in equation (14) is the same as that in equation (13). The restriction (12) can then be derived in the same fashion.

### 4 Empirical results

The proposed restricted four-factor GATSM presents a parsimonious Gaussian Affine Term Structure Model (GATSM) incorporating both cross-sectional information in yields and that in expected excess returns. The restricted four-factor model can be summarized by the following two equations:

\[
X_{t+1} = \mu + \rho X_t + \Sigma \varepsilon_{t+1}, \varepsilon_{t+1} \sim N(0, I),
\]

\[
Y_t^o = A + BX_t + e_t, e_t \sim N(0, \Omega).
\]

The factor is \( X_t = \left( P' \ H' \right)' \). Respectively, \( \mu, \rho \) and \( \Sigma \varepsilon_{t+1} \) are constant, autoregressive coefficient and shock in the factor’s VAR(1) process. The vector of observed yields with maturities ranging from one to five years is denoted by \( Y_t^o \), i.e. \( Y_t^o = \left( y_{t}^{(12)} \ y_{t}^{(24)} \ y_{t}^{(36)} \ y_{t}^{(48)} \ y_{t}^{(60)} \right)' \). Matrices \( A \) and \( B \) are obtained by stacking corresponding \( a_n \)'s and \( b_n \)'s as follows

\[
A = \left( a_{12} \ a_{24} \ a_{36} \ a_{48} \ a_{60} \right)',
\]

\[
B = \left( b_{12} \ b_{24} \ b_{36} \ b_{48} \ b_{60} \right)'. \quad (15)
\]
Measurement error is represented by $e_t$ and is assumed to follow a normal distribution with mean zero and covariance $\Omega$. I assume that $\Omega$ is diagonal following the literature. Let $\Theta$ collect all parameters, i.e. $\Theta = (\mu, \rho, \mu^Q, \rho^Q, \delta_0, \delta_1, \Sigma, \Omega)$. The log likelihood is given by

$$
\mathcal{L} (Y_T, Y_{T-1}, ..., Y_1, X_T, X_{T-1}, ... X_2 | X_1; \Theta) = \\
- \frac{(T - 1) \times 4}{2} \log (2\pi) - \frac{T - 1}{2} \log (|\Sigma\Sigma'|) - \frac{1}{2} \sum_{t=2}^{T} (X_t - \mu - \rho X_{t-1})' (\Sigma\Sigma')^{-1} (X_t - \mu - \rho X_{t-1}) \\
- \frac{T \times 5}{2} \log (2\pi) - \frac{T}{2} \log (|\Omega|) - \frac{1}{2} \sum_{t=2}^{T} (Y_t^o - A - BX_t)' \Omega^{-1} (Y_t^o - A - BX_t) .
$$

(16)

Note that $A$ and $B$ only depend on subsets of $\Theta$, i.e. $A = A (\delta_0, \delta_1, \mu^Q, \rho^Q, \Sigma)$ and $B = B (\delta_1, \rho^Q)$.

**Concentrated MLE** Before estimating the model by MLE, I demean both $Y_t^o$ and $X_t$. Instead of estimating $a_n$s and $\mu$, I use sample averages. Consequentially, $\Sigma$ does not enter the last line of equation [16]. Then estimation can be done by concentrated MLE, which makes optimization algorithms significantly faster to converge. By using sample averages, restrictions between $a_n$s imposed by no-arbitrage are removed. While we may lose efficiency, we decrease the risk of misspecifying restrictions. Also, [Hamilton and Wu (forthcoming)] showed that $a_n$s from the model are poorly estimated, which provides another justification for demeaning. The log likelihood function with demeaned variables can be written as follows:

$$
\mathcal{L} (\overline{Y}_t^o, \overline{Y}_{t-1}^o, \overline{X}_t, \overline{X}_{t-1}, ... | \overline{X}_1; \rho^Q, \delta_1, \Sigma, \Omega) = \\
- \frac{(T - 1) \times 4}{2} \log (2\pi) - \frac{T - 1}{2} \log (|\Sigma\Sigma'|) - \frac{1}{2} \sum_{t=2}^{T} (\overline{X}_t - \rho \overline{X}_{t-1})' (\Sigma\Sigma')^{-1} (\overline{X}_t - \rho \overline{X}_{t-1}) \\
- \frac{T \times 5}{2} \log (2\pi) - \frac{T}{2} \log (|\Omega|) - \frac{1}{2} \sum_{t=1}^{T} (\overline{Y}_t^o - B \overline{X}_t)' \Omega^{-1} (\overline{Y}_t^o - B \overline{X}_t) ,
$$

where $\overline{Y}_t^o$ and $\overline{X}_t$ are demeaned yields and factors. As we can see, given any $\rho$, $\rho^Q$ and $\delta_1$, the values of $\hat{\Omega}$ and $\hat{\Sigma}\hat{\Sigma}'$ that maximize the log likelihood function would just be the sample
covariance matrix of residuals, i.e.

$$\bar{\Sigma}\bar{\Sigma}'(\rho, \rho^Q, \delta_1) = \frac{1}{T-1} \sum_{t=2}^{T} (\bar{X}_t - \rho \bar{X}_{t-1})(\bar{X}_t - \rho \bar{X}_{t-1})',$$  \hspace{1cm} (17)

$$\hat{\Omega}(\rho, \rho^Q, \delta_1) = \text{diag} \left( \frac{1}{T} \sum_{t=1}^{T} (Y_t - BX_t) \odot (Y_t - BX_t) \right),$$ \hspace{1cm} (18)

where $\odot$ is element-by-element multiplication and $\text{diag}(.)$ is a diagonal matrix using input vector as its diagonal. The corresponding log likelihood function would take the value

$$-\frac{(T-1) \times 4}{2} \log (2\pi) - \frac{T-1}{2} \left( \log |\bar{\Sigma}\bar{\Sigma}'| \right) - \frac{(T-1) \times 4}{2} - \frac{T \times 5}{2} \log (2\pi) - \frac{T}{2} \left( \log |\hat{\Omega}| \right) - \frac{T \times 5}{2}. \hspace{1cm} (19)$$

Thus maximizing the log likelihood function is equivalent to finding $\hat{\rho}, \hat{\rho}^Q$ and $\hat{\delta}_1$ so that expression (19) is maximized with expressions for $\hat{\Omega}$ and $\bar{\Sigma}\bar{\Sigma}'$ in equations (17) and (18) with $B$ given by equations (15) and (6) subject to the further restrictions derived in Section 3, i.e. $\hat{\delta}_{1,H} = 0, \hat{\rho}^Q_{PH} = 0$ and $((\hat{\rho}^Q)^m)_{PP} = (\hat{\rho}^m)_{PP}$. Note that $\rho^Q$ is identified up to $(\rho^Q)^{12}$ since only yields of one through five years are used. $\Sigma$ is assumed to be lower triangular for identification purpose. The optimization problem is solved by the MATLAB function “fminunc” with 500 different initial values.
Estimates  The estimated $\rho$ with robust standard errors (see Hamilton(1994) p.145) in parentheses is given below:

$$
\hat{\rho} = 
\begin{pmatrix}
1.0058 & -0.3166 & 1.0722 & 0.2217 \\
0.0053 & 0.0298 & 0.2286 & 0.0164 \\
0.0036 & 0.9237 & 0.5234 & 0.0217 \\
(0.0006) & (0.0059) & (0.0344) & (0.0054) \\
-0.0013 & 0.0059 & 0.8709 & -0.0031 \\
(0.0002) & (0.0021) & (0.0134) & (0.0020) \\
-0.0065 & 0.1849 & -0.5476 & 0.7840 \\
(0.0077) & (0.0460) & (0.3216) & (0.0217)
\end{pmatrix}.
$$

The first two elements of $\hat{\rho}_{PH}$, i.e. the first two elements of the last column of $\hat{\rho}$, are significantly different from zero (with t-statistic larger than 1.96). It follows that the forecasting factor has significant impacts on the dynamics of level and slope under the $P$ measure. Therefore, the conventional GATSM only including the first three principal components misspecifies the dynamics of yields. Recall that $\hat{\rho}_{PH}^Q = $ is a zero vector and $\Sigma \lambda_1 = \rho - \rho^Q$. Thus the statistical significance of $\hat{\rho}_{PH}$ means that both level and slope risk are priced by the forecasting factor. This finding differs from restrictions imposed in Cochrane and Piazzesi(2009) and Duffee(2011) that only price for level risk is time-varying, but agrees with Joslin, Priebsch, and Singleton(2010). Estimates of other parameters are not of direct interest and thus not shown.

I also estimate the conventional three-factor GATSM via concentrated MLE after demeaning variables. Further computation simplification can be achieved for the unrestricted three-factor model by recognizing that $\rho$ can be directly estimated with OLS as Joslin, Singleton, and Zhu(2011) pointed out.
5 Implications for the term premium

In this section, I assess economic significance of the model by examining its implications. In particular, I focus on employing the model to decompose the yield curve. The yield curve can be decomposed into two components, expected future yields and term premia. While the former depends on markets’ projection of future monetary policy, the latter reflects compensations for bearing interest rate risk. Identifying each component’s contribution is of great interest to both market practitioners and monetary authorities. Theoretical studies, Campbell and Cochrane(1999), Bansal and Yaron(2004) and Wachter(2006) for examples, mainly support countercyclical term premia, and they present a challenge for the conventional GATSM. The model generates almost acyclical term premia, which is pointed by Bauer, Rudebusch, and Wu(2012) among others and is reproduced in Figure 6. In contrast to the conventional GATSM, which use the level, slope and curvature both to explain contemporaneous yields and to predict future yields, my restricted four-factor GATSM provides an extra factor to forecast future yields. Because decomposing the yield curve hinges on how to form expectations of future yields, one possible source accounting for the failure of the conventional GATSM to generate countercyclical term premia could be inadequate factor specification. This is plausible because the term premia estimated by Kim and Wright(2005) were countercyclical when they filtered out information not only from the cross-section of yields but also the survey data on expected interest rates. Another possible explanation is the small-sample bias argued by Bauer, Rudebusch, and Wu(2012). In the remaining of the section, I decompose the yield curve with my model and compare the result with that of the conventional three-factor GATSM.

The term premium for a \( n \)-period bond is defined as follows

\[
TP_t^{(n)} = y_t^{(n)} - \frac{1}{n/12} E_t \left( y_{t+12}^{(n)} + \ldots + y_{t+n-12}^{(n)} \right).
\]

(21)

This is the excess return that investors required as compensation for holding a long term
bond instead of a series of short ones. Since expected yields can not be observed, different models lead to distinct term premia measures. With estimates obtained in Section 4 I can compute the term premia for 2- through 5-year bonds implied by both the restricted four-factor GATSM and the conventional three-factor GATSM. They are displayed in Figure 4. Not surprisingly, the two models disagree on their term premia measures because they have different specifications for the dynamics of yields. Since economic theories suggest term premia are countercyclical, it is worthwhile to examine how these two measures change over the business cycle. In Figures 5 and 6 I plot the term premia with the growth rate of the Industrial Production Index (IPI) for the two models, respectively. The risk premia from the restricted four-factor GATSM display marked cyclical variation with their values rising significantly during recessions, consistent with economic theories. The correlations between the term premia and the IPI growth rate are around $-0.15$. On the other hand, the term premia from the conventional three-factor GATSM are almost acyclical. Their correlations with the IPI growth rate are around $-0.04$. This comparison suggests that the inability of the conventional three-factor GATSM to generate countercyclical risk premia may originate from factor misspecification, i.e. only using factors that span the cross-section of yields and omitting factors that drive their dynamics.

6 Conclusion

I propose a parsimonious GATSM exploiting information in expected excess returns, which has been ignored by commonly used GATSMs. I construct a single forecasting factor to summarize the cross-section of expected excess returns by reduced rank forecasting regression with a large predictor set. This factor can explain realized excess returns with $R^2$ up to 0.71 and subsumes the role of the tent-shaped combination of forward rates in Cochrane and Piazzesi(2005) in in-sample prediction. I include the forecasting factor along with the commonly used level, slope and curvature in a GATSM, and I impose restrictions considering
the fact that the forecasting factor and the three principal components span the cross-section of expected returns and the cross-section of yields, respectively. In contrast to the traditional three-factor (the level, slope and curvature) GATSM, the resulting restricted four-factor GATSM generates countercyclical term premia as economic theories suggest.
7 Appendix

Derivation of bond pricing in the GATSM

Under no-arbitrage restrictions, an asset with payoff $g(X_{t+1})$ is priced by

$$\text{Price} (X_t) = E_t (\exp (m_{t+1}) g (X_{t+1})).$$

Plug in equations (1) and (2),

$$E_t (\exp (m_{t+1}) g (X_{t+1})) = \exp (-r_t) \exp \left(-\frac{1}{2} \lambda_t' \lambda_t \right) E_t \left(\exp \left(-\lambda_t' \Sigma^{-1} (X_{t+1} - \mu_t)\right) g (X_{t+1})\right),$$

where $\mu_t \equiv \mu + \rho X_t$. Since $X_{t+1}$ follows a conditional normal distribution with conditional mean $\mu_t$ and conditional variance $\Sigma \Sigma'$,

$$E_t \left(\exp \left(-\lambda_t' \Sigma^{-1} (X_{t+1} - \mu_t)\right) g (X_{t+1})\right) = \exp \left(\frac{1}{2} \lambda_t' \lambda_t \right) (2\pi)^{-K/2} |\Sigma \Sigma'|^{-1/2} \int \exp \left(-\frac{1}{2} \left(X_{t+1} - \mu_t\right)' (\Sigma \Sigma')^{-1} \left(X_{t+1} - \mu_t\right)\right) g (X_{t+1}) dX_{t+1},$$

where $\mu_t^Q = \mu_t - \Sigma \lambda_t$. Plug in equation (3),

$$\mu_t^Q = \underbrace{(\mu - \Sigma \lambda_0)}_{\mu^Q} + \underbrace{(\rho - \Sigma \lambda_1)}_{\rho^Q} X_t.$$

Therefore,

$$\text{Price} (X_t) = E_t (\exp (m_{t+1}) g (X_{t+1}))$$

$$= \exp (-r_t) (2\pi)^{-K/2} |\Sigma \Sigma'|^{-1/2} \int \exp \left(-\frac{1}{2} \left(X_{t+1} - \mu_t^Q\right)' (\Sigma \Sigma')^{-1} \left(X_{t+1} - \mu_t^Q\right)\right) g (X_{t+1}) dX_{t+1},$$

$$= \exp (-r_t) E_t^Q (g (X_{t+1})).$$
Assume \( y_t^{(n)} = \frac{a_n}{n} + \frac{b'_n}{n} X_t \). From \( \exp \left( p_t^{(n)} \right) = \exp ( -r_t ) E^Q_t \left( \exp \left( p_{t+1}^{(n-1)} \right) \right) \), we have

\[
\exp ( -a_n - b'_n X_t ) = \exp ( -\delta_0 - \delta'_1 X_t ) E^Q_t \left( \exp \left( -a_{n-1} - b'_{n-1} X_{t+1} \right) \right).
\]

Since \( X_{t+1} \) is conditionally normally distributed with mean \( \mu^Q + \rho^Q X_t \) and variance \( \Sigma \Sigma' \) under the Q measure,

\[
E_t^Q \left( \exp \left( -b'_{n-1} X_{t+1} \right) \right) = \exp \left( -b'_{n-1} \left( \mu^Q + \rho^Q X_t \right) + \frac{1}{2} b'_{n-1} \Sigma \Sigma' b_{n-1} \right).
\]

Matching constant terms and coefficients in front of \( X_t \) on both sides of the equation, we can derive the recursive formulation for \( a_n, b_n \) for \( n \geq 2 \):

\[
a_n = \delta_0 + a_{n-1} + b'_{n-1} \mu^Q - \frac{1}{2} b'_{n-1} \Sigma \Sigma' b_{n-1},
\]

\[
b'_n = \delta_1 + b'_{n-1} \rho^Q.
\]

The recursion starts from \( a_1 = \delta_0, b'_1 = \delta'_1 \) by equation \( \boxed{4} \).
8 Figures

Figure 1: Realized and predicted excess returns

Note: Blue lines are realized excess returns. Red lines are predicted excess returns from forecasting regressions specified by equation 8. Shaded areas are recession periods. Four panels are for different maturities.
Figure 2: Forecasting regression coefficients on forward rates

Note: The plot presents estimated coefficients from regression of the average excess return on forward rates specified by equation 9.

Figure 3: Time series for factors

Note: Four panels plot time series for level, slope, curvature, and the forecasting factor, respectively. Shaded areas are recession periods.
Figure 4: Term premia: the restricted four-factor GATSM v.s. the conventional three-factor GATSM

Note: Term premia defined in equation 21 for 2- through 5-year bonds are displayed in four panels. Blue lines are computed with the restricted four-factor GATSM and green lines are computed with the conventional three-factor GATSM. Shaded areas are recession periods.
Figure 5: Term premia and the IPI growth rate: the restricted four-factor GATSM

Note: Standardized 12-month moving average for term premia from the restricted four-factor GATSM (blue lines) and for the IPI growth rate (red lines) are plotted. Shaded areas are recession periods. Four panels correspond to four different maturities.
Figure 6: Term premia and the IPI growth rate: the conventional three-factor GATSM

Note: Standardized 12-month moving average for term premia from the conventional three-factor GATSM (green lines) and for the IPI growth rate (red lines) are plotted. Shaded areas are recession periods. Four panels correspond to four different maturities.
References


Wu, Jing Cynthia, and Fan Dora Xia (2013) “Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound” .