CIRJE-F-784

How Does Yield Curve Predict GDP Growth? A Macro-Finance Approach Revisited

Junko Koeda
University of Tokyo

January 2011

How Does Yield Curve Predict GDP Growth? A Macro-Finance Approach Revisited

October 2010
Revised: January 2011

Junko Koeda [1 2]

Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo.

Abstract

This note analyzes the yield-curve predictability for GDP growth by modifying the time-series property of the interest rate process in Ang, Piazzesi, and Wei (2006). When interest rates have a unit root and term spreads are stationary, the short rate's forecasting role changes, and the combined information from the short rate and term spread intuitively reveals the relationship between the shift of yield curves and GDP growth.

Key words: JEL Classification: C13, C32, E43, E44, E27
Term structure of interest rates, GDP growth, Unit root, Estimation

1 Introduction

Yield-curve variables, particularly those that capture the level and slope of yield curves measured by short-term interest rates and term spreads, are recognized as useful leading indicators of GDP growth (surveyed in Stock and Watson (2003)). Which of these variables can best predict GDP growth? While this question has typically been examined using unconstrained regressions, it may be better answered using constrained regressions based on a macro-finance term-structure model where bond yields and macro variables are jointly modeled with cross-sectional restrictions relating short- and long-term interest rates.

[1] Tel: +81-3-5841-5649. Email address: jkoeda@e.u-tokyo.ac.jp.
[2] I am grateful to seminar participants at the Bank of Japan, Keio University, and the University of Tokyo for their useful discussions and comments. I acknowledge funding from the Nomura Foundation for Academic Promotion. All remaining errors are my own.

To this end, some researchers (Ang, Piazzesi, and Wei (2006), APW henceforth [3]) propose adopting a macro-finance, no-arbitrage term structure model, where the cross-sectional restrictions arise from the absence of arbitrage in bond markets. APW use an affine term structure framework in which the yields are expressed as the affine function of yield factors. They include the short rate and term spread as yield factors so that their estimated results are comparable with those based on unconstrained linear regressions that include the two factors in the regressors. Their model provides better out-of-sample GDP forecasts, especially on the longer-horizon, than single-equation unconstrained regression models.

APW's estimated results, however, include several other distinguishing features: (i) short rates forecast GDP growth better than spreads and the spread has little additional predictive power, although the existing work typically supports the predictive power of the spread (e.g., Stock and Watson (2003), Bordo and Haubrich (2008)), and (ii) an increase in the short rate is associated with negative GDP growth, ceteris paribus, although policymakers often see that a bear steepening of yield curves (an increase in both the short rate and term spread) is associated with positive growth.

These artifacts may arise from the underlying assumption of the time-series property of the short rate. APW assume that the level of the short rate is linear stationary, even though the augmented Dickey-Fuller unit-root test cannot reject the null that it has a unit root (for details of such evidence, see Appendix A). If the true short-rate process is difference stationary, the long-term effects of interest rate movements can differ considerably.

To better understand the mechanisms of the APW model, this note modifies the original APW model by considering Campbell and Shiller (1987, CS henceforth) type state dynamics in which the short rate has a unit root while the term spread is stationary (i.e., the short and long rates are cointegrated). [4] [5] In contrast to the original APW findings, the estimated results of this note

[3] For a variant of APW that addresses term-spread predictability, see for example, Ichiue (2004). This author extends APW by adding inflation to the state variable, along with more lags. While the link between GDP growth and inflation is important, I exclude inflation from the model, in the same spirit as in APW, to make the results directly comparable to the literature, which uses only term structure information.

[4] A unit root process could induce disputable long-run implications. For example, the I(1) specification implies that forward rates fall without limit as maturity increases (Campbell et al. (1997)). In this paper, I limit my analysis to bond yield data up to a 5-year maturity.

[5] An alternative approach to address the persistency of the short-term rate is to consider a stationary, non-linear process (e.g., Kozicki and Tinsley (2001)). However, such a process is often at odds with the affine framework.

support the predictability of both the short rate and term spread on GDP growth. Moreover, the short rate's role in forecasting GDP growth changes. In the original model, an increase in the short rate is associated with negative growth. In the modified model, on the other hand, bear steepening is associated with positive growth, while bear flattening produces a mixed result, albeit the short rate increases in either case. Lastly, an out-of-sample forecasting exercise indicates that the modified model has higher predictive accuracy than the original model in the investigated period.

This note proceeds as follows. Section 2 describes the model. Section 3 explains the estimation strategy. Section 4 presents the estimated and forecasting results. Finally, Section 5 concludes.

2 The Model

2.1 The state dynamics

The APW's state dynamics is originally the VAR(1) of the level of the short rate (r), 20-quarter yield spread ($r^{20} - r$), and GDP growth rate (g), which is given by

$$X_{t+1} = \Gamma_0 + \Gamma X_t + \Sigma \varepsilon_{t+1}, \text{ where } X_t = \left[r_t, r_t^{20} - r_t, g_t \right]', \tag{1}$$

where Γ_0 is a 3×1 vector, Γ is a 3×3 matrix, Σ is a 3×3 lower triangular matrix, and ε is a 3×1 random shock vector assumed to be standard normal and independent of each other and over time.

In this note, I replace the above state dynamics with a CS-type VAR(1) of the *first difference* in the short rate (Δr), 20-quarter yield spread, and GDP growth rate. I do this because the augmented Dickey-Fuller test cannot reject the null of unit root in the short rate (for details see Appendix A). The state dynamics is given by

$$\tilde{X}_{t+1} = \Gamma_0 + \Gamma \tilde{X}_t + \Sigma \varepsilon_{t+1}, \text{ where } \tilde{X}_t = \left[\Delta r_t, r_t^{20} - r_t, g_t \right]', \tag{2}$$

(2) can be rewritten into the following "level" dynamics,

$$
\begin{aligned}
X_{t+1} &= \Gamma_0 + \Gamma_1 X_t + \Gamma_2 X_{t-1} + \Sigma\varepsilon_{t+1} \qquad (3)\\
\text{where } X_t &= \left[r_t, r_t^{20} - r_t, g_t\right]', K = \begin{bmatrix} 1\ 0\ 0 \\ 0\ 0\ 0 \\ 0\ 0\ 0 \end{bmatrix},\\
\Gamma_1 &= K + \Gamma, \quad \Gamma_2 = -\Gamma K. \qquad (4)
\end{aligned}
$$

This is a restricted VAR(2) with restrictions imposed on Γ_1 and Γ_2 in (4). Thus, our state dynamics is a two-order system, albeit with the restrictions given by (4).

2.2 Pricing kernel and the price of risk

As in APW, suppose that the pricing kernel is given by

$$
\begin{aligned}
m_{t+1} &= \exp(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1}), \qquad (5)\\
\lambda_t &= \lambda_0 + \lambda_1 X_t + \lambda_2 X_{t-1}, \qquad (6)
\end{aligned}
$$

where the price of risk takes an affine form in state variables X_t and X_{t-1}, as handled in many existing affine term structure models. λ_0 is a 3×1 vector and λ_1 and λ_2 are 3×3 matrices. Then the log price of a n-period bond takes the following affine form

$$
\begin{aligned}
p_t^n &= \exp\left(\bar{A}_n + \bar{\mathbf{B}}_n X_t + \bar{\mathbf{C}}_n X_{t-1}\right),\\
\text{where } \bar{A}_{n+1} &= \bar{A}_n + \bar{\mathbf{B}}_n \Gamma_0 - \bar{\mathbf{B}}_n \Sigma\lambda_0 + \frac{1}{2}\bar{\mathbf{B}}_n \Sigma\Sigma'\bar{\mathbf{B}}_n', \qquad (7)\\
\bar{\mathbf{B}}_{n+1} &= \bar{\mathbf{B}}_n\left(\Gamma_1 - \Sigma\lambda_1\right) + \bar{\mathbf{C}}_n - e_1', \qquad (8)\\
\bar{\mathbf{C}}_{n+1} &= \bar{\mathbf{B}}_n\left(\Gamma_2 - \Sigma\lambda_2\right). \qquad (9)
\end{aligned}
$$

where e_i is a 3×1 vector of zeros with a 1 in the ith element. The derivation of these recursive equations is given in Appendix B.

The n-period, zero-coupon bond yield is given by

$$
\begin{aligned}
r_t^n &= A_n + \mathbf{B}_n X_t + \mathbf{C}_n X_{t-1}, \qquad (10)\\
\text{where } A_n &= -\bar{A}_n/n,\ B_n = -\bar{\mathbf{B}}_n/n,\ C_n = -\bar{\mathbf{C}}_n/n.
\end{aligned}
$$

For the 20-quarter yield equation ($n = 20$), the following two implications of the model need to be consistent with each other: (i) the model-implied 20-quarter bond yield is given by Eq. (10) with $n = 20$, (ii) the 20-quarter yield is the sum of the first two factors in X_t. To ensure this consistency, as in APW, the yield coefficients for the 20-quarter yield spread must be

$$A_{20} = 0, \mathbf{B}_{20} = [1, 1, 0], \mathbf{C}_{20} = [0, 0, 0]. \tag{11}$$

2.3 Term spread stationarity

The unit-root assumption on the short rate implies that longer maturity yields also have a unit root; however it is widely acknowledged that the two yields are cointegrated. Against this background, I rewrite the yield equation (Eq. (10)) as the following yield-spread equation,

$$r_t^n - r_t = A_n + \mathbf{F}_n X_t + \mathbf{C}_n X_{t-1}, \tag{12}$$
$$\text{where } \mathbf{F}_n \equiv \mathbf{B}_n - e_1'. \tag{13}$$

The following proposition provides a condition under which the term spread in equation (Eq. (12)) is stationary.

Proposition 1 *Yield spread is stationary, for any given Σ, if and only if $(\lambda_1 + \lambda_2)e_1' = 0$.*

See Appendix C for proof.

3 Estimation Strategy

Although the model dynamics is a second-order system, it is effectively a first-order system if the state is taken to be $\tilde{X}$. Accordingly, I restrict λ_t to be an affine function of $\tilde{X}_t$, so that the λ_t is now given by

$$\begin{aligned} \lambda_t &= \lambda_0 + \lambda_1 X_t + \lambda_2 X_{t-1}, \\ &= \lambda_0 + \Lambda \tilde{X}_t, \end{aligned} \tag{14}$$
$$\text{where } \lambda_1 = \Lambda, \quad \lambda_2 = -\Lambda K, \text{and } \Lambda \text{ is a } 3 \times 3 \text{ matrix.} \tag{15}$$

This restriction is stronger than that of Proposition 1[6] and would save the number of prices-of-risk parameters to be estimated. The yield-spread equation (12) can now be simplified as

$$r_t^n - r_t = A_n + \mathbf{F}_n \tilde{X}_t, \tag{16}$$

[6] I do not impose (15) in Section 2 to illustrate the condition needed only for yield-spread stationarity.

where $\mathbf{F}_n X_t + \mathbf{C}_n X_{t-1}$ can be rewritten as $\mathbf{F}_n \tilde{X}_t$, by Eqs. (4), (9), (15), and (C.3). The corresponding system of equations can be expressed with the dynamics of $\tilde{X}$ with the yield-spread equations appended to $\tilde{X}$. The system will be stationary if yield spreads are stationary because the dynamics of $\tilde{X}$ is stationary.

For illustration purposes, I use the same data sources and sample period as APW. That is, I use quarterly data on interest rates and GDP growth rate from 1964Q1 to 2001Q4. The 1-quarter zero coupon bond yields are used for the short-term interest rate and zero-coupon bond yields of 4-, 8-, 12-, 16-, and 20-quarter maturities are used for longer maturities; These bond yields are from the CRSP US Treasury Database (the Fama-Bliss Discount Bond Files for 4-, 8-, 12-, 16-, and 20-quarter data and from the Risk-Free Rate Files for 1-quarter data). All bond yields are continuously compounded and expressed at a quarterly frequency in percent. Real GDP growth data are taken from the FRED database and expressed in quarterly percent changes.

The system of equations to be estimated can be summarized as follows:

$$\underbrace{\begin{pmatrix} \tilde{X}_{t+1} \\ R_t - r_t \end{pmatrix}}_{\equiv Y_{t+1}} = \underbrace{\begin{pmatrix} \Gamma_0 \\ A \end{pmatrix}}_{\equiv A_Y} + \underbrace{\begin{pmatrix} \Gamma \\ \mathbf{F} \end{pmatrix}}_{\equiv \mathbf{F}_Y} \tilde{X}_t + \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma^m \end{pmatrix} \begin{pmatrix} \varepsilon_{t+1} \\ \varepsilon^m_{t+1} \end{pmatrix}, \tag{17}$$

$$\text{where } R_t = \left[r_t^4, r_t^8, r_t^{12}, r_t^{16} \right]', A = [A_4, A_8, A_{12}, A_{16}]', \mathbf{F} = [\mathbf{F}_4, \mathbf{F}_8, \mathbf{F}_{12}, \mathbf{F}_{16}]',$$

where R_t is a 4×1 vector of bond yields with maturities indicated by the superscript numbers (in quarters). The yield dynamics is an affine function of the state variables with a 4×1 coefficient vector A and a 4×3 matrix $\mathbf{F}$ corresponding to the constant term and $\tilde{X}$ respectively. A and $\mathbf{F}$ are time invariant with maturities indicated by the subscript numbers. Their elements are derived from the recursive equations; in other words, the model implicitly imposes cross-equation restrictions reducing the number of parameters to be estimated. Measurement errors ε^m are assumed to have constant variance and Σ^m is a diagonal matrix. We estimate this system using the constrained maximum likelihood method in which the constraints are $A_{20} = 0$ and $\mathbf{F}_{20} = [0, 1, 0]$ (for details see Appendix D).[7] [8]

[7] That is, the sample here is $(y_1, ..., y_T) = (X_1, R_0 - r_0; X_2, R_1 - r_1; ..., X_T, R_{T-1} - r_{T-1})$. It would be more natural to consider the sample $(X_1, R_1 - r_1; X_2, R_2 - r_2; ..., X_T, R_T - r_T)$, but the usual factorization argument can be more readily applied to the former. If the sample size T is large, the choice of the sample would not matter for the point estimation.

[8] For the standard errors under the constrained maximization, I calculated the consistent estimates of the asymptotic variance discussed in equation (7.4.22) in

4 Estimated and Forecasting Results

4.1 Estimation summary

The parameter estimates of the model are reported in Table 1. The positive coefficient in the last row of the Γ matrix corresponding to term spread (0.893) is consistent with the common view that a rise in term spread predicts growth. The diagonal elements in the λ_1 matrix are statistically significant indicating that the prices of risk are driven by movement of the corresponding factor.

4.2 The implied growth regression

As mentioned above, the predictability of yield-curve variables has been widely examined via unconstrained growth regressions. This subsection, following the approach discussed by APW, links the estimated parameters to those in the typical growth regression model estimated in the literature.

The growth regression model is given by

$$g_{t\to t+k} = \alpha_k + \beta_k' \tilde{X}_t + \varepsilon_{t+k,k}, \tag{18}$$

where $g_{t\to t+k}$ is the average quarterly GDP growth rate from period t to $t+k$, α_k and β_k are the k-period forecast coefficients and $\varepsilon_{t+k,k}$ is the corresponding error term. The coefficient β_k equals that for $\tilde{X}_t$ in the least-square projection of $g_{t\to t+k}$ on constant and $\tilde{X}_t$, thus [9]

$$\beta_k = \left[V(\tilde{X}_t)\right]^{-1} Cov(\tilde{X}_t, g_{t\to t+k}). \tag{19}$$

The coefficient vector β_k takes different values depending on the underlying models. First, β_k based on the single-equation regression model of $g_{t\to t+k}$ on a constant and $\tilde{X}$ can be estimated via OLS. Second, β_k based on the unrestricted VAR(1) of $\tilde{X}$ can be obtained by computing the right-hand side of (19) using the unrestricted VAR parameters. Lastly, β_k based on the modified APW, i.e., the constrained VAR(1) of $\tilde{X}$ imposing no-arbitrage equations, can be obtained by computing the right-hand side of (19) using the estimated model parameters reported in Table 1.

Hayashi (2000).

[9] $V(\tilde{X}_t)$ is the unconditional variance of $\tilde{X}$ with $vec(V(\tilde{X})) = (I-\Gamma\otimes\Gamma)^{-1} vec(\Sigma\Sigma')$, and $Cov(\tilde{X}_t, g_{t\to t+k})$ is the covariance of $\tilde{X}$ and $g_{t\to t+k}$ given by $V(\tilde{X})\tilde{\Gamma}_k'\Gamma' e_3$ with $\tilde{\Gamma}_k = (I-\Gamma)^{-1}(I-\Gamma^k)/k$.

Table 2 reports the β_k implied by the modified model (left panel) and those reported in the original model (right panel) for $k = 1, 4, 8, 12$. The corresponding standard errors are calculated via the delta method. There are several messages in these results. First, the two models have contrasting results for term spread predictability on growth: none of the term-spread coefficients implied by the original model are statistically significant, whereas those implied by the modified model are significant for all forecasting horizons. Moreover, the size of the latter coefficients is notably larger. This means that the term spread plays a nontrivial role in predicting GDP growth under the modified model. Second, while the short rate contains important information about GDP growth in either model, its economic role differs. In the original model, an increase in the short rate is always associated with negative growth, ceteris paribus. On the other hand, in the modified model, the combined information from the short-rate (in first difference) and the term-spread terms suggests that bear steepening is associated with positive growth, and bull flattening (i.e., both the short rate and term spread decrease) is associated with negative growth. Bear flattening or bull steepening induces a mixed result depending on the degree of flattening (steepening) of yield curves relative to changes in the short rate. Lastly, the growth coefficients are insignificant on the longer-horizons in either model, suggesting that forward-looking financial variables play a more important role in forecasting GDP growth than current growth rate.

4.3 Out-of-sample forecast

I perform out-of-sample forecasting over the period of 2002Q1-2006Q4 with the forecasting ranging from 1- to 12-quarters.[10] This forecasting exercise is rolling. The left panel of Table 3 reports the root mean square error (RMSE) ratios of the models in comparison (i.e., original APW, the single-equation regression model, and the unrestricted VAR(1) of $\tilde{X}$) to the modified APW. The forecast error is defined by taking the actual $g_{t\to t+k}$ subtracting the corresponding forecast. The RMSE ratios suggest that the modified APW performs better than the original APW on the longer-horizon (i.e., 4, 8, and 12 quarter horizons). Further, they indicate that the model has higher predictive accuracy than the single-equation regression model on the medium-horizon (i.e., 4 and 8 quarter horizons), despite the fact that the model has many more parameters. Lastly, they show little difference in forecasting performance between the modified APW (i.e., a restricted VAR(1) of $\tilde{X}$) and the unrestricted VAR.

The predictive accuracy can also be checked using formal hypothesis testing. The right panel of Table 3 reports the modified Diebold-Mariano (MDM) test

[10] Thus, to obtain the 12-quarter forecast error at 2006Q4, for example, I use the data up to 2009Q4, including the recent global financial crisis.

statistics proposed by Harvey et al. (1998) comparing the modified APW with the model in comparison.[11] The differential loss is based on the mean-square errors, and the forecast error is defined by taking the actual k-period ahead quarterly growth rate (g_{t+k}) subtracting the corresponding forecast. A significantly negative statistic indicates that the modified APW performs better than the model in comparison. The MDM test confirms the earlier results that the modified APW performs better than the original APW, particularly on the medium-term horizons, and its predictive accuracy is similar to that of the unrestricted VAR.

These forecasting results, however, may be subject to the choice of out-of-sample period. For example, if the period covered begins in the late-80s, the predictability of the yield-curve model could deteriorate due to a decline in the volatility of macroeconomic variables, as is often argued in the literature.

5 Conclusion

This note analyzes the yield-curve predictability for GDP growth using a modified APW model. In line with the original APW findings, the estimated results support that the macro-finance, no-arbitrage, term-structure model improves predictive accuracy for GDP growth especially on the medium horizon, compared with the unconstrained single-equation model. In the modified model, however, the short rate plays a different forecasting role, and together with the term spread, it intuitively reveals the relationship between the shift of yield curves and GDP growth. Further, the modified model forecasts GDP growth better than the original model in the investigated period.

The theoretical literature has yet to reach a consensus on how exactly yield curves predict GDP growth. While the model suggests that bear steepening signals economic recovery near the end of recession, other factors may influence it, for example, the fear of rising fiscal debt, as discussed recently in policy circles. In future research, fiscal risks should be incorporated in the model, and their links with other macro variables, for example, expected inflation, should be addressed.

[11] I do not use the Clark and West (2007) test because the modified APW is not nested in the original APW or the unrestricted VAR in the Clark and West's sense.

References

[1] Ang, A., Piazzesi, M., Wei, M., 2006. What does the yield curve tell us about GDP growth? Journal of Econometrics, 131, 359-403.

[2] Bordo, M.D., Haubrich, J.G., 2008. Forecasting with the yield curve; level, slope, and output 1875-1997. Economics Letters, 99, 48-50.

[3] Campbell, J. Y., Lo, A. W., Mackinlay, A.C., 1997. The Econometrics of Financial Markets, Princeton University Press, Princeton, New Jersey.

[4] Campbell, J.Y., Shiller, R. J., 1987. Cointegration and tests of present value models. Journal of Political Economy, Vol. 95, No. 5 (Oct., 1987), pp. 1062-1088.

[5] Clark, T., West, K., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics, 138, 291-311.

[6] Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics, 13, 253-263.

[7] Harvey, D.I., Leybourne, S.J., Newbold, P., 1998. Tests for forecast encompassing. Journal of Business and Economic Statistics 16, 254–259.

[8] Hayashi, F, 2000. Econometrics, Princeton University Press December, 2000.

[9] Ichiue, H., 2004. Why Can the Yield Curve Predict Output Growth, Inflation, and Interest Rates? An Analysis with an Affine Term Structure Model. Bank of Japan Working Paper Series, July 2004.

[10] Kozicki, S., Tinsley, P.A., 2001. Shifting endpoints in the term structure of interest rates. Journal of Monetary Economics 47, 613-652.

[11] Stock, J.H., Watson, M. W., 2003. Forecasting output and inflation: the role of asset prices. Journal of Economic Literature, Vol. 41, No. 3 (Sep., 2003), pp. 788-829.

A Unit root and cointegration tests on bond yields

In this appendix, I carry out the augmented Dickey-Fuller test on bond yields and term spreads with various maturities. Table A.1. shows that, for all maturities, the null hypothesis that the levels of the yields have a unit root cannot be rejected even at the 10-percent level, but the null that the first differences of the yields have a unit root can be rejected; the null hypothesis that the spreads over the short rate have a unit root can be rejected at the 1 percent level. In short, the test indicates that all considered bond yields have a unit root, however, the longer-rates are cointegrated with the short rate.

B Derivation of the recursive equations

I can confirm that the n-period bond pricing formula in

$$\begin{aligned} p_t^{n+1} &= E_t\left(m_{t+1}p_{t+1}^n\right) \\ &= E_t\left[\begin{array}{c} \exp(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1}) \\ \times\exp(\bar{A}_n + \bar{\mathbf{B}}_n X_{t+1} + \bar{\mathbf{C}}_n X_t) \end{array}\right] \\ &= \exp(-r_t + \bar{A}_n - \frac{1}{2}\lambda_t'\lambda_t) \\ &\times E_t\left[\exp(-\lambda_t'\varepsilon_{t+1} + \bar{\mathbf{B}}_n X_{t+1} + \bar{\mathbf{C}}_n X_t)\right]. \end{aligned}$$

Plugging in the dynamics of (3) into above gives

$$\begin{aligned} p_t^{n+1} =& \exp(-r_t + \bar{A}_n - \frac{1}{2}\lambda_t'\lambda_t) \\ &\times E_t\left[\exp(-\lambda_t'\varepsilon_{t+1} + \bar{\mathbf{B}}_n X_{t+1} + \bar{\mathbf{C}}_n X_t)\right] \\ =& \exp(-r_t + \bar{A}_n - \frac{1}{2}\lambda_t'\lambda_t) \\ &\times E_t\left[\exp\left(-\lambda_t'\varepsilon_{t+1} + \bar{\mathbf{B}}_n\left(\begin{array}{c}\Gamma_0 + \Gamma_1 X_t \\ +\Gamma_2 X_{t-1} + \Sigma\varepsilon_{t+1}\end{array}\right) + \bar{\mathbf{C}}_n X_t\right)\right] \\ =& \exp(-r_t + \bar{A}_n - \frac{1}{2}\lambda_t'\lambda_t + \bar{\mathbf{B}}_n\Gamma_0 + \left(\bar{\mathbf{B}}_n\Gamma_1 + \bar{\mathbf{C}}_n\right)X_t + \bar{\mathbf{B}}_n\Gamma_2 X_{t-1}) \\ &\times E_t\left[\exp(\left(-\lambda_t' + \bar{\mathbf{B}}_n\Sigma\right)\varepsilon_{t+1})\right] \\ =& \exp\left(\begin{array}{c} -r_t + \bar{A}_n + \bar{\mathbf{B}}_n\Gamma_0 + \frac{1}{2}\bar{\mathbf{B}}_n\Sigma\Sigma'\bar{\mathbf{B}}_n' \\ +\left(\bar{\mathbf{B}}_n\Gamma_1 + \bar{\mathbf{C}}_n\right)X_t + \bar{\mathbf{B}}_n\Gamma_2 X_{t-1} - \bar{\mathbf{B}}_n\Sigma\left(\lambda_0 + \lambda_1 X_t + \lambda_2 X_{t-1}\right)\end{array}\right). \end{aligned}$$

The last equality relies on ε_t being i.i.d. standard normal and the dynamics of λ_t given by (14).

The bond price equation can finally be rewritten as

$$\begin{aligned} p_t^{n+1} =& E_t\left(m_{t+1}p_{t+1}^n\right) \\ =& \exp\left(\begin{array}{c} \underbrace{\bar{A}_n + \bar{\mathbf{B}}_n\Gamma_0 - \bar{\mathbf{B}}_n\Sigma\lambda_0 + \frac{1}{2}\bar{\mathbf{B}}_n\Sigma\Sigma'\bar{\mathbf{B}}_n'}_{\equiv\bar{A}_{n+1}} \\ +\underbrace{\left(\bar{\mathbf{B}}_n\left(\Gamma_1 - \Sigma\lambda_1\right) + \bar{\mathbf{C}}_n - e_1'\right)}_{\equiv\bar{\mathbf{B}}_{n+1}}X_t + \underbrace{\bar{\mathbf{B}}_n\left(\Gamma_2 - \Sigma\lambda_2\right)}_{\equiv\bar{\mathbf{C}}_{n+1}}X_{t-1}\end{array}\right). \end{aligned}$$

C Proof of proposition 1

$\mathbf{F}_n X_t$ can be rewritten as

$$\begin{aligned}\mathbf{F}_n X_t &= \mathbf{F}_n \left(K X_t + (I-K) X_t\right) \\ &= \mathbf{F}_n e_1 r_t + \mathbf{F}_n \begin{bmatrix} 0 \\ r^{20}_{t-1} - r_{t-1} \\ g_{t-1} \end{bmatrix}. \end{aligned} \tag{C.1}$$

Similarly, $\mathbf{C}_n X_{t-1}$ can be rewritten as

$$\mathbf{C}_n X_{t-1} = \mathbf{C}_n e_1 r_{t-1} + \mathbf{C}_n \begin{bmatrix} 0 \\ r^{20}_{t-1} - r_{t-1} \\ g_{t-1} \end{bmatrix}. \tag{C.2}$$

The second RHS terms in (C.1) and (C.2) are stationary so that,

$$\mathbf{F}_n X_t + \mathbf{C}_n X_{t-1} = [\mathbf{F}_n e_1, \mathbf{C}_n e_1] \begin{bmatrix} r_t \\ r_{t-1} \end{bmatrix} + \text{stationary components.}$$

For stationarity of the first RHS term, the first elements of $\mathbf{F}_n$ and $\mathbf{C}_n$ must add up to zero, i.e.,

$$(\mathbf{F}_n + \mathbf{C}_n) e_1 = 0, \text{ for all } n \tag{C.3}$$

By (8), (9), and (13), $(\mathbf{F}_{n+1} + \mathbf{C}_{n+1}) e_1$ can be expressed as

$$\begin{aligned}(\mathbf{F}_{n+1} + \mathbf{C}_{n+1}) e_1 &= \frac{n}{1+n} \begin{bmatrix} (\mathbf{F}_n + e_1') (\Gamma_1 - \Sigma\lambda_1) e_1 + \mathbf{C}_n e_1 - e_1' e_1 ... \\ ... + (\mathbf{F}_n + e_1') (\Gamma_2 - \Sigma\lambda_2) e_1 \end{bmatrix}, \\ &= \frac{n}{1+n} \begin{bmatrix} (\mathbf{F}_n + e_1') (\Gamma_1 + \Gamma_2) e_1 + \mathbf{C}_n e_1 - 1 ... \\ ... - (\mathbf{F}_n + e_1') \Sigma (\lambda_1 + \lambda_2) e_1 \end{bmatrix}, \\ &= \frac{n}{1+n} \begin{bmatrix} \mathbf{F}_n e_1 + 1 + \mathbf{C}_n e_1 - 1 ... \\ ... - (\mathbf{F}_n + e_1') \Sigma (\lambda_1 + \lambda_2) e_1 \end{bmatrix}, \end{aligned} \tag{C.4}$$

The equality in (C.4) holds because

$$
\begin{aligned}
(\mathbf{F}_n + e_1')\,(\Gamma_1 + \Gamma_2)\,e_1 &= (\mathbf{F}_n + e_1')\,(K + \Gamma - \Gamma K)\,e_1, \\
&= (\mathbf{F}_n + e_1')\,(K + \Gamma(I - K))\,e_1, \\
&= (\mathbf{F}_n + e_1')\,K e_1, \\
&= (\mathbf{F}_n + e_1')\,e_1, \\
&= \mathbf{F}_n e_1 + 1.
\end{aligned}
$$

Note that for any given Σ, the last term in (C.4) (i.e., $(\mathbf{F}_n + e_1')\,\Sigma\,(\lambda_1 + \lambda_2)\,e_1$) disappears if and only if,

$$
(\lambda_1 + \lambda_2)e_1 = [0, 0, 0]', \tag{C.5}
$$

so that

$$
(\mathbf{F}_{n+1} + \mathbf{C}_{n+1})\,e_1 = \frac{n}{1+n}\,(\mathbf{F}_n + \mathbf{C}_n)\,e_1.
$$

Now, via mathematical induction, it can be shown that (C.3) holds with (C.5). That is, for $n = 1$,

$$
(\mathbf{F}_1 + \mathbf{C}_1)e_1 = 0,
$$

where $\mathbf{F}_1 = [0, 0, 0]$ and $\mathbf{C}_1 = [0, 0, 0]$. For $n = 2$,

$$
\begin{aligned}
(\mathbf{F}_2 + \mathbf{C}_2)e_1 &= \frac{2}{3}(\mathbf{F}_1 + \mathbf{C}_1)e_1, \\
&= 0.
\end{aligned}
$$

QED.

D The log likelihood function

We estimate the model dynamics by numerically maximizing the following log-likelihood function:

$$
L(\theta) = -\frac{1}{2}\sum_{t=1}^{T} \log(\det(\Omega)) - \frac{1}{2}\sum_{t=1}^{T} u_t \Omega^{-1} u_t,
$$

subject to the constraints on the recursive equations (i.e., $A_{20} = 0$ and $\mathbf{F}_{20} = [0, 1, 0]$), where θ is the vector of parameters to be estimated;

$$\theta = [\Gamma_0, \Gamma, \Sigma, \Sigma^m, \lambda_0, \Lambda],$$

$$\Omega = \begin{pmatrix} \Sigma\Sigma' & 0 \\ 0 & \Sigma^m\Sigma^{m\prime} \end{pmatrix},$$

Ω is the covariance-variance matrix, and u is defined by

$$u_t = Y_t - A_Y - \mathbf{F}_Y \tilde{X}_{t-1}.$$

State dynamics

	Γ_0	Γ			Σ		
ΔShort rate	-0.045	-0.186	0.055	0.023	0.284	--	--
	(0.035)	(0.090)	(0.119)	(0.041)	(0.011)	--	--
Spread	0.063	0.084	0.814	-0.008	-0.164	0.121	--
	(0.027)	(0.063)	(0.091)	(0.027)	(0.019)	(0.008)	--
GDP growth	0.461	0.721	0.893	0.164	0.206	0.148	0.773
	(0.124)	(0.360)	(0.379)	(0.097)	(0.099)	(0.088)	(0.041)

Risk premia

	λ_0	λ_1			λ_2		
Short rate	-0.481	-0.845	-0.659	0.095	0.845	--	--
	(0.155)	(0.305)	(0.405)	(0.145)	(0.305)	--	--
Spread	-1.727	--	-0.994	0.015	--	--	--
	(0.329)	--	(0.471)	(0.131)	--	--	--
GDP growth	-8.527	--	0.912	-1.099	--	--	--
	(4.059)	--	(2.491)	(0.334)	--	--	--

Measurement error

3 months	12 months	36 months	60 months
0.085	0.063	0.042	0.031
(0.009)	(0.008)	(0.005)	(0.001)

Table 1. Estimated coefficients. This table reports estimated coefficients. Numbers in parentheses indicate standard errors. Prices of risk parameters with large standard errors (i.e., the second and third elements in the first column of the λ_1 matrix) are set to zero. Measurement error is the standard deviation of error corresponding to each maturity.

horizon	Modified APW coefficients			Original APW coefficients		
qts	Short rate in first difference	Spread	Growth	Short rate	Spread	Growth
1	0.72	0.89	0.16	-0.27	0.32	0.26
	(0.36)	(0.38)	(0.10)	(0.14)	(0.33)	(0.09)
4	0.23	0.81	0.05	-0.28	0.25	0.08
	(0.09)	(0.33)	(0.04)	(0.14)	(0.31)	(0.04)
8	0.13	0.60	0.02	-0.25	0.14	0.04
	(0.05)	(0.24)	(0.02)	(0.14)	(0.25)	(0.03)
12	0.09	0.45	0.01	-0.22	0.08	0.02
	(0.03)	(0.19)	(0.01)	(0.13)	(0.20)	(0.02)

Table 2. Modified and original APW coefficients. This table reports predictive coefficients implied by the modified (left panel) and original (right panel) APW models. Numbers in parentheses indicate standard errors.

horizon	RMSE ratios			MDM test	
qts	Original APW	Single-eq.	VAR	Original APW	VAR
1	1.09	0.99	0.99	-1.01	1.32
4	1.20	1.07	0.99	-2.34**	1.39
8	1.52	1.18	0.99	-1.72*	1.65
12	1.28	0.99	1.01	0.61	0.87

Table 3. Out-of-sample forecasts. The left panel reports RMSE ratios relative the modified APW and the right panel reports the MDM test statistics. * and ** indicate the 10 and 5 percent significant levels respectively. The out-of-sample period is 2002Q1-2006Q4.

	t statistics		
qts	level	first difference	spread over 1-qt yield
1	-2.03	-6.79	---
4	-1.57	-16.43	-8.66
8	-1.89	-15.78	-6.27
12	-1.80	-15.51	-4.47
16	-1.76	-15.01	-4.10
20	-1.73	-14.95	-4.12

Table A.1. Unit root test on bond yields. This table reports the *t*-statistics with the null hypothesis that the level of (the second column) or the first difference (the third column) of each bond yield has a unit root. The maturities of the bond are reported in the first column. Note that the critical values for 1- and 10-percent levels are -3.47 and -2.58 respectively. Lag lengths are selected based on the Schwarz Information Criterion.